Some Optimal and Sequential Experimental Designs with Potential Applications to Nanostructure Synthesis and Beyond

Professor Tirthankar Dasgupta **Li Zhu**

Some Optimal and Sequential Experimental Designs with Potential Applications to Nanostructure Synthesis and Beyond

Abstract

Design of Experiments (DOE) is an important topic in statistics. Efficient experimentation can help an investigator to extract maximum information from a dataset. In recent times, DOE has found new and challenging applications in science, engineering and technology. In this thesis, two different experimental design problems, motivated by the need for modeling the growth of nanowires are studied.

In the first problem, we consider issues of determining an optimal experimental design for estimation of parameters of a complex curve characterizing nanowire growth that is partially exponential and partially linear. A locally D-optimal design for the non-linear change-point growth model is obtained by using a geometric approach. Further, a Bayesian sequential algorithm is proposed for obtaining the D-optimal design. The advantages of the proposed algorithm over traditional approaches adopted in recent nano-experiments are demonstrated using Monte-Carlo simulations.

The second problem deals with generating space-filling design in feasible regions of complex response surfaces with unknown constraints. Two different types of sequential design strategies are proposed with the objective of generating a sequence of design points that will quickly carve out the (unknown) infeasible regions and generate more and more points in the (unknown) feasible region. The generated design is

space-filling (in certain sense) within the feasible region. The first strategy is model independent, whereas the second one is model-based. Theoretical properties of proposed strategies are derived and simulation studies are conducted to evaluate the performance of proposed strategies. The strategies are developed assuming that the response function is deterministic, and extensions are proposed for random response functions.

Contents

Acknowledgments

No person deserves my thanks more than Prof. Tirthankar Dasgupta. With his enthusiasm, inspiration, and great efforts, I have learned and enjoyed statistics during my five years here. Without his constant inspirations and help, I could never finish the challenging PhD study. Prof. Dasgupta not only offered me careful guidance in statistics but also helped me a lot during my life. There is an old Chinese adage "he who teaches me for one day will be my mentor for life". I believe the guidance I have obtained from Prof. Dasgupta will be extremely beneficial in my future life and career. I feel greatly honored to be his first PhD student.

I also want to express my most sincere gratitude to my thesis committee members, Prof. Xiao-li Meng and Carl Morris. I want to thank them for their generous help in serving as my committee members. Moreover, Prof. Xiao-li Meng inspired me in interesting research topic in astro-statistics. Prof. Morris also offered me great help in Stat 210/211, which introduced me into advanced statistics in my first year. Moreover, I also want to thank Mr. Steven Finch, who proofread my thesis and gave me very useful advice.

I also want to show my sincere gratitude to all great professors with whom I worked. Prof. Jeff Wu and Roshan Joseph have given me very useful comments and ideas on my thesis work. Prof. Qiang Huang has inspired my work in nanotechnology and provided the data for my thesis. I want to thank Prof. Charles Nunn, who gave me fascinating introduction on statistics in evolutionary biology and the work with him is very interesting and inspiring. I also want to thank other faculty members within statistics department, Prof. Jun Liu , Samuel Kou, Donald Rubin, Joseph Blitzstein, Yoonjung Lee and Natesh Pillai, their interesting courses and seminars

has helped me a lot during my study and research.

I am also very grateful to the staffs in statistics department: Betsey Cogswell, James Matejek, Kevin Rader, Dale Rinkel, Maureen Stanton and Ellen Weene, who assisted me in teaching, preparing post-qualifying talks, picking up mails etc.

I also want to thank my friends and colleagues in statistics department, Chao Du, Simeng Han, Bo Jiang, Zhan Li, Bronwyn Loong, Sergiy Nesterko, Thomas Xiao Tong, Xianchao Xie, Xiaojin Xu, Amy Yuan, Yuan Yuan etc. Their accompany made the five years journey fun and interesting. I also want to thank all members in Harvard Poker Club, which introduced me statistics in a different way.

Last but most importantly, I wish to thank my parents for their consistent help and support during my PhD study. Without their love, support and assistance, I can never arrive at this stage.

To Wenting, who knew what she did for me.

Chapter 1

Introduction

1.1 Introduction to experimental design

Design of Experiments (DOE) is one of the most important topics in statistics. Experimentation, which allows a researcher to investigate the outcomes of a system when the input variables are *purposely* changed, is widely used in different fields. DOE can help the researcher to extract maximum information from experimentation and obtain the optimal setting of input, an extremely important task due to the large cost of modern day experimentation.

DOE originated from agriculture between 1918 and the 1940s, when Ronald A. Fisher proposed and used experiments for different agricultural activities. Since the experimentation in agriculture is usually large in scale, and taking long time to complete (Wu and Hamada, 2000), different methods such as blocking, randomization and replication were proposed. Different design methods, such as fractional factorial designs, were also proposed. During World War II, DOE has been widely used in

chemical and process industries. The ideas of optimal design and sequential design were developed during this time. After that, during the second industrial era between the late 1970s and 1990s, DOE was developed in quality improvement, where Genichi Taguchi proposed robust parameter design. After the 1990s, DOE attracted more research attention and has been widely used in many new fields.

DOE can help an investigator to extract maximum information from data for decision making. DOE is especially important where data are expensive and difficult to gather, which is common in modern day experimentation. DOE has been widely used in different fields. For example, DOE has been used in process modeling and optimization in industries. In agriculture, DOE can be used to determine whether to utilize new agricultural technologies to the farm. DOE also has wide usage in quality control and other fields. Moreover, DOE has attracted recent research interest in many newly developed fields, such as nanoscience, computer modeling and complex systems.

1.2 Problems addressed in thesis

In this thesis, two different experimental design problems are explored. The first problem involves investigation of an optimal experimental strategy to estimate the parameters of an exponential-linear change-point model. The second problem involves investigation of a strategy for obtaining a space-filling design in a feasible sub-space of the design space, where the feasibility constraints are unknown. We explore two different strategies to address the latter problem. One is a naive strategy with proven theoretical properties and the other is based on a Gaussian process model.

Both of these problems are motivated by applications in nanotechnology. In Section 1.3, we explain the specific challenges that motivate these research problems, and also how these problems are connected.

However, it is worthwhile to note that, in spite of being motivated by applications in nanotechnology, the second problem is more general and has potential applications in computer experiments, Bayesian computation and mechanical engineering. We will discuss some of these applications in Section 1.4.

1.3 Motivating examples from nanotechnology

Nanotechnology is the study of structures on an atomic and molecular scale. Research in nanotechnology deals with structures with at least one dimension ranging from 1 to 100 nanometres (one nanometer = 10^{-9} meter). Nanotechnology has been one of the most popular research topics in the 21^{st} century and can provide fundamental understanding of different materials, which is widely used in different fields. Among different types of nanostructures, nanowires have received significant attention among physicists and material scientists.

A topic of considerable research interest is the study of conditions which facilitate the growth of nanowires. A schematic diagram of such an experimental study is given in Figure 1.1. The black-box model represents the unknown response model. The input factors, typically called 'treatment factors' or 'control factors' in the experimental design literature, are denoted by the vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_p)$. Growth temperature, pressure, gas flow rate, etc. are examples of such factors.

For each setting of $\boldsymbol{X}$, a 'growth curve' representing the growth of nanowires over

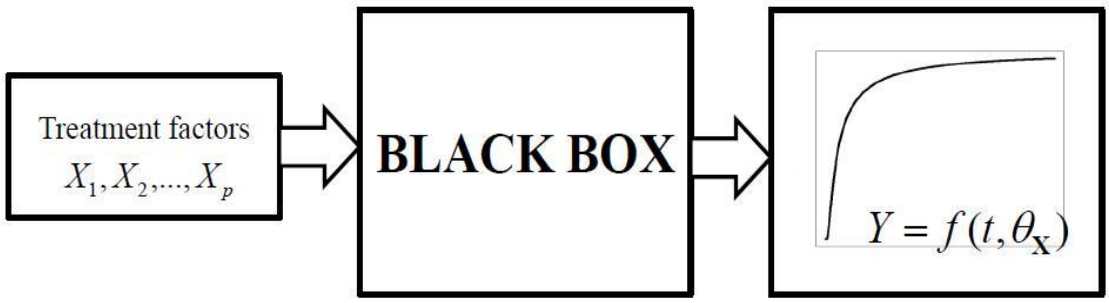

Figure 1.1: Schematic diagram of experimental study in nanotechnology.

time can be observed. The growth, represented by the length or weight of nanowires, is a function of time, and is of the form:

$$Y = f(t, \theta_{\boldsymbol{X}}) \tag{1.1}$$

where $\theta_{\boldsymbol{X}}$ is a vector of parameters that depends on $\boldsymbol{X}$. Broadly speaking, the experimenter has the following objectives:

1. For a given setting of $\boldsymbol{X}$, identifying a suitable model and estimating the model parameters.

2. Conducting experiments at different settings of $\boldsymbol{X}$ to ascertain which among the variables $X_1, \cdots, X_p$ affect the parameter vector $\theta_{\boldsymbol{X}}$ and how.

With respect to the first objective, we consider a scenario where the functional

form $f(t, \theta_{\boldsymbol{X}})$ is known from physical knowledge (and verified by earlier experiments). The experimental design question is, for a given setting of $\boldsymbol{X}$, how should the experimenter choose the time points $t_1, \cdots, t_n$ so that the parameters $\theta_{\boldsymbol{X}}$ can be estimated most efficiently.

In order to achieve the second objective, experiments under different control variables $X_1, \cdots, X_p$ should be conducted to identify the model and model parameters as a function of the treatment factors. If the factors are categorical and/or the response is simple (such as linear or quadratic), a factorial experiment can be used to conduct experimentation for all possible combinations of factors. However, when factors are continuous, the response function is complex and the number of factors is large, a factorial experiment is not appropriate. Then a space-filling design which can provide exploration over different combinations of controlled variables could be used. Examples of such strategies include Latin hypercube design (LHD) (Santner et al., 2003) and uniform design (Fang and Lin, 2000).

However, one major challenge associated with nanowires synthesis that makes experimentation for the second task difficult is the huge variation in morphology (the study of how the shape and form of molecules affect their chemical properties) over different regions of the design space. There are frequently *large*, *irregular* and *unknown* regions of no-morphology (referred as infeasible region). For example, under some control variables, the growth is extremely low or even zero. Experiments at such region are useless for the model identification and parameters estimation.

Due to the cost of experimentation, existing space-filling design cannot be directly used since it would waste large amount of resources in the infeasible region. Instead,

a novel *space-filling strategy with unknown constraints* which can: (i) avoid the *large*, *irregular* and *unknown* infeasible region and (ii) generate space-filling design within the feasible region leading to precise estimation of model and model parameters should be proposed.

1.4 Why the second problem is more general: Examples of applications from computer and mechanical engineering

A strategy which can address the second problem mentioned in Section 1.3 has wide potential usage in computer experiment and mechanical engineering. Space-filling designs have been exclusively used to model deterministic outputs from computer experiments (e.g. Sacks et al., 1989). Many such experiments involve known or unknown constraints. For example, Stinstra et al. (2003) give an example where a simulation scheme in the design optimization process of Philips television tubes should be identified. However, 44 non-box constraints on simulating parameters result from geometric restriction on the tube. Consequently, simulation schemes failing to satisfy the constraints cannot be used. Another example comes from sheet metal spinning in mechanical engineering (Henkenjohann et al., 2005), where highly complex input-output relationships combined with a large number of constraints are involved. Although some work has been done on such problems where constraints are known, little work has been done when the constraints are unknown. The space-filling strategy with unknown constraints from Section 1.3 could then possibly be used.

1.5 Structure of thesis

The rest of thesis is organized as follows. In Chapter 2, an optimal design for exponential-linear change-point growth curve is proposed to address the first problem in Section 1.3. In Chapters 3, 4 and 5, space-filling strategies with unknown constraints are developed to address the second problem. Specifically, in Chapter 4, a naive space-filling strategy is proposed and its theoretical properties are investigated. In Chapter 5, a design strategy based on Gaussian process modeling is proposed. Possible extensions to the two strategies are discussed in Chapter 6.

Chapter 2

A D-optimal design for estimation of parameters of an exponential-linear growth curve of nanostructures

2.1 Introduction

Nanostructured materials and processes have been estimated to increase their market impact to about $340 billion per year in the next 10 years. However, high cost of processing has been a major barrier in transferring the fast-developing nanotechnology from laboratories to industry applications. The process yield of current nano devices being very poor (typically 10% or less), there is a need of process improvement methodologies in nanomanufacturing. For this purpose, it is important

to understand the growth mechanisms of nanostructures better through controlled experimentation. Statistical design of experiments are therefore expected to play an important role (Dasgupta et al., 2008, Lu et al., 2009) in this area.

Let $Y(t)$ denote the value of a quality characteristic (e.g., weight of nanostructures grown, see Huang et al., 2011) of a nanostructure synthesized using the vapor-liquid-solid (VLS) process at time t at a *specific* location. Huang et al. (2011) experimentally investigated six weight kinetics models to study weight change of nanostructures over time. In this paper we shall consider three of these models which are of the form

$$Y(t) = g_{\boldsymbol{\theta}}(t) + \epsilon, \quad t > 0, \tag{2.1}$$

where $g_{\boldsymbol{\theta}}(\cdot)$ is a parametric function of time t, $\boldsymbol{\theta}$ represents a parameter vector, and the error terms ϵ are independently and identically distributed as $N(0, \sigma^2)$. The three models differ with respect to the specification for $g_{\boldsymbol{\theta}}(\cdot)$. The first model has a pure exponential growth function where

$$g_{\boldsymbol{\theta}}(t) = \alpha_1 e^{-\alpha_2/t}, \quad t > 0. \tag{2.2}$$

In all future discussions, we shall refer to model (2.1) with $g_{\boldsymbol{\theta}}(\cdot)$ defined by (2.2) as model M_1. The other two models, henceforth referred to as models M_3 and M_4 respectively (model M_2 will be introduced later), are characterized by an exponential-

linear change-point growth function:

$$g_{\boldsymbol{\theta}}(t) = \begin{cases} \alpha_1 e^{-\frac{\alpha_2}{t}}, & t_{\min} \leq t < t_0 \\ \\ a + bt, & t_0 \leq t \leq t_{\max}, \end{cases}, \tag{2.3}$$

where $\alpha_1 > 0, \alpha_2 > 0$, $b > 0$; $t_{\min}$ and $t_{\max}$ denote respectively the (known) earliest and latest time points at which it is feasible to conduct a trial; and t_0 is an *unknown time* such that the length $L(t_0)$ of a nanostructure at time t_0 is the diffusion length defined in growth kinetics. Models M_3 and M_4 differ by the fact that the former has a strong assumption of continuity and differentiability of $g_{\boldsymbol{\theta}}(t)$ at t_0, whereas the latter relaxes the assumption of differentiability and only assumes continuity at t_0. It can easily be seen (Section 2.2) that models M_3 and M_4 can be reduced to three and four parameter models respectively.

Based on their experimental results, Huang et al. (2011) identified model M_4 as the most appropriate model among the three. They used an experimental strategy (i.e., selection of time points $t \in [t_{\min}, t_{\max}]$ at which growth was observed) which utilized the prior knowledge that the change-point t_0 was more likely to occur early on in the growth process. Thus, instead of selecting equispaced points within $t_{\min} = 0.5$ minutes and $t_{\max} = 210$ minutes, they chose 19 time points that were more or less equally-spaced on a log-scale. The basic experiment was replicated three times.

However, this experimental strategy of conducting 57 experimental runs to estimate the parameters of the model was considered too expensive (gold is usually used as a catalyst) and time consuming (terminating the process at a certain time t, taking the substrate out of the furnace, and measuring the weight using a high-precision

microbalance). Naturally, scientists are interested in determining with an *efficient* experimentation strategy that would: (i) allow available scientific knowledge to be incorporated into the experimental design, (ii) involve as few trials as possible, and yet (iii) allow precise estimation of model parameters. In this article, we derive a locally D-optimal experimental design for the above estimation problem using a geometric approach and also propose a sequential Bayesian strategy that converges to the locally D-optimal design at the true parameter values.

A locally D-optimal design for model M_1 (in a slightly different form) was derived by Mukhopadhyay and Haines (1995). In this paper we shall derive locally D-optimal designs for models M_3 and M_4. To do this, we will introduce another model M_2, which has the same functional form of $g_{\boldsymbol{\theta}}(\cdot)$ as M_3 and M_4 given by (2.3), but assumes that the change point t_0 is *known*. Note that model M_2 is not of any practical interest; we introduce it merely to explain the geometrical argument with a better logical flow that starts with a two-dimensional problem.

There has been some recent work on Bayesian D-optimal design with change points (Atherton et al., 2009). However, our problem does not have the typical "on-line" nature that is associated with the problem considered by Atherton et al. (2009) and most change-point problems. In our problem, an experimental unit (a substrate on which nanostructures are grown) can generate only a single data point at the time it is taken out of the furnace and examined under the microscope. Consequently, the assumptions used in Atherton et al. (2009) (pre-specified distance d between any two design points and no replicated observations) do not hold for the current problem. There has also been some recent work on D-optimal designs for spline

models (Biedermann et al., 2009).

This Chapter is organized as follows. In the following section, we give an overview of locally D-optimal designs and state the work done for model M_1. In Section 2.3, we describe the geometric approach for obtaining D-optimal designs, and use it to derive locally D-optimal designs for models M_2 and M_3. Since a D-optimal design for a non-linear model like (2.3) will involve the true values of the parameters, in Section 2.4 we suggest a sequential Bayesian strategy to obtain designs that are expected to converge to the true D-optimal design. In Section 2.5, we re-visit the experimental strategy adopted by Huang et al. (2011) for estimation of parameters of model M_3 and demonstrate how the proposed strategy can help reduce the number of experimental runs. Section 2.6 presents some concluding remarks and opportunities for future work.

2.2 Information matrix and locally D-optimal designs

As mentioned in Section 2.1, the objective is to design an experiment that will help us estimate the parameter vector $\boldsymbol{\theta}$ with a reasonably high degree of precision. We need to determine at which time points t the trials should be conducted. Mathematically speaking, the design space $\mathcal{X}$ is the closed interval $[t_{min}, t_{max}]$. There is a given σ–field $\mathcal{F}$ of sets in $\mathcal{X}$, containing all one-point sets, and a design measure ξ, which is a probability measure on $(\mathcal{X}, \mathcal{F})$. Let Ξ denote the class of all such design measures. Finding an optimal design means selecting a measure ξ that will optimize a certain criterion associated with the Fisher Information Matrix $\mathbf{I}(\boldsymbol{\theta}, \xi)$ for the model

under consideration. Specifically, here we consider a D-optimal design (**?**Federov, 1972), which is obtained by maximizing the determinant of $\mathbf{I}(\boldsymbol{\theta}, \xi)$ over the possible designs ξ. In other words, the optimal design measure ξ^* will be such that

$$\det \int \mathbf{I}(\boldsymbol{\theta}, \xi^*)\xi^*(dt) \geq \det \int \mathbf{I}(\boldsymbol{\theta}, \xi)\xi(dt), \tag{2.4}$$

for all $\xi \in \Xi$.

We now derive the information matrices for the models M_1 through M_4.

Information matrix for model M_1:

For model M_1, the Fisher information matrix at a single design point t is:

$$\mathbf{I}(\boldsymbol{\theta}, \xi_t) = \frac{1}{\sigma^2} e^{-2\alpha_2/t} \begin{bmatrix} 1 & -\frac{\alpha_1}{t} \\ -\frac{\alpha_1}{t} & \frac{\alpha_1^2}{t^2} \end{bmatrix},$$

where ξ_t denotes a one-point probability measure with support t and σ^2 is dispersion parameter. The above matrix can be expressed in the form $\frac{1}{\sigma^2}\mathbf{v}_1\mathbf{v}_1'$, where

$$\mathbf{v}_1' = e^{-\frac{\alpha_2}{t}}(1, -\frac{\alpha_1}{t}). \tag{2.5}$$

Information matrices for models M_2 and M_3:

For models M_2 and M_3, using the continuity and differentiability of $g_{\boldsymbol{\theta}}(t)$ at $t = t_0$, it is easy to see that

$$a = \alpha_1(1 - \frac{\alpha_2}{t_0})e^{-\frac{\alpha_2}{t_0}}, \text{ and} \tag{2.6}$$

$$b = \frac{\alpha_1\alpha_2}{t_0^2}e^{-\frac{\alpha_2}{t_0}}, \tag{2.7}$$

so that model M_2 is essentially a two-parameter model with $\boldsymbol{\theta} = (\alpha_1, \alpha_2)$, whereas model M_3 has three unknown parameters α_1, α_2 and t_0.

For model M_2, the Fisher information matrix can be expressed in the form $\frac{1}{\sigma^2}\mathbf{v}_2\mathbf{v}_2'$, where

$$\mathbf{v}_2' = \begin{cases} e^{-\frac{\alpha_2}{t}}(1, -\frac{\alpha_1}{t}), & t < t_0 \\ e^{-\frac{\alpha_2}{t_0}}\left(1 - \frac{\alpha_2}{t_0} + \frac{\alpha_2 t}{t_0^2}, -\frac{\alpha_1}{t_0}(2 - \frac{\alpha_2}{t_0} - \frac{t}{t_0} + \frac{\alpha_2 t}{t_0^2})\right). & t \geq t_0 \end{cases} \tag{2.8}$$

Similarly, the information matrix for model M_3 can be expressed as $\frac{1}{\sigma^2}\mathbf{v}_3\mathbf{v}_3'$, where

$$\mathbf{v}_3' = \begin{cases} e^{-\frac{\alpha_2}{t}}\left(1, -\frac{\alpha_1}{t}, 0\right), & t < t_0 \\ e^{-\frac{\alpha_2}{t_0}}\left(1 - \frac{\alpha_2}{t_0} + \frac{\alpha_2 t}{t_0^2}, -\frac{\alpha_1}{t_0}(2 - \frac{\alpha_2}{t_0} - \frac{t}{t_0} + \frac{\alpha_2 t}{t_0^2}), \frac{\alpha_1\alpha_2}{t_0^2}\left(2 - \frac{\alpha_2}{t_0} - \frac{2t}{t_0} + \frac{\alpha_2 t}{t_0^2}\right)\right). & t \geq t_0 \end{cases} \tag{2.9}$$

Information matrix for model M_4:

Since for model M_4 the function $g_{\boldsymbol{\theta}}(t)$ is no longer differentiable at t_0, we have an extra slope parameter b. The intercept a of the linear part can be obtained from

$$a = \alpha_1 e^{-\frac{\alpha_2}{t_0}} - bt_0. \tag{2.10}$$

In this model, the information matrix can be expressed in the form $\frac{1}{\sigma^2}\mathbf{v}_4\mathbf{v}_4'$, where

$$\mathbf{v}_4' = \begin{cases} e^{-\frac{\alpha_2}{t}}(1, -\frac{\alpha_1}{t}, 0, 0), & t < t_0 \\ \left(e^{-\frac{\alpha_2}{t_0}}, -\frac{\alpha_1}{t_0}e^{-\frac{\alpha_2}{t_0}}, \frac{\alpha_1\alpha_2}{t_0^2}e^{-\frac{\alpha_2}{t_0}} - b, t - t_0\right). & t \geq t_0 \end{cases} \tag{2.11}$$

Note that in this case, $\mathbf{v}_4$ is no longer continuous at time t_0.

The above representation of the information matrix $\mathbf{I}$ in the form $\mathbf{v}\mathbf{v}'$ will be seen

to be very useful in derivation of the locally D-optimal design as well as the sequential design. Note that since models M_1 to M_4 are intrinsically non-linear, the Fisher information matrix in all three cases depends on the model parameters. Therefore the optimal design ξ^* that maximizes the D-optimality criterion $\det \int \mathbf{I}(\boldsymbol{\theta}, \xi)\xi(dt)$ will actually be a *locally* D-optimal design (Chernoff, 1953) at $\boldsymbol{\theta}$. Moreover, since dispersion parameter σ^2 serves as scalar variable for information matrices for M_1 to M_4, it will not affect the optimization to obtain the locally D-optimal design.

Proceeding in lines with the proof of Theorem 1 in Mukhopadhyay and Haines (1995), the following Theorem can be established.

Theorem 1. *The locally D-optimal design for model M_1 is a balanced two points design with unique support at the points t_1^*, t_2^*, where*

$$t_2^* = t_{\max},$$

and

$$t_1^* = max(\frac{\alpha_2 t_2^*}{\alpha_2 + t_2^*}, t_{\min}).$$

Proof. Consider the information matrix $I(\theta, \xi_t)$ for a single design time t. Then the set $\{1, e^{-2\alpha_2/t}, \frac{1}{t}e^{-2\alpha_2/t}\}$ forms a Tchebycheff system on any fixed interval. Then it follows from the results of Karlin and Studden (1966, pg. 333) and of Federov (1972, pg. 85-86) that, for any fixed parameters, the D-optimal design is based on exactly two points of support.

As observed by Mukhopadhyay and Haines (1995), a necessary and sufficient con-

dition for a design ξ to be locally D-optimal is

$$d_{\boldsymbol{\theta}}(\xi, t) = tr[\mathbf{I}(\boldsymbol{\theta}, \xi_t)\mathbf{I}(\boldsymbol{\theta}, \xi)^{-1}] - 2 \leq 0, \tag{2.12}$$

for all t in the design space.

For any design point t in the design space, it is easy to see that

$$d_{\boldsymbol{\theta}}(\xi, t) = 2e^{-\frac{2\alpha_2}{t}} \frac{e^{-\frac{2\alpha_2}{t_1}}(\frac{1}{t} - \frac{1}{t_1})^2 + e^{-\frac{2\alpha_2}{t_2}}(\frac{1}{t} - \frac{1}{t_2})^2}{e^{-2\alpha_2(\frac{1}{t_1}+\frac{1}{t_2})}(\frac{1}{t_2} - \frac{1}{t_1})^2}, \tag{2.13}$$

and we want $d_{\boldsymbol{\theta}}(\xi, t) \leq 0$ for all points between t_{min} and t_{max}. Taking derivative with respect to t, we have

$$d'_{\boldsymbol{\theta}}(\xi, t) = \frac{4e^{-\frac{\alpha_2}{t}}}{t^2 e^{-2\alpha_2(\frac{1}{t_1}+\frac{1}{t_2})}(\frac{1}{t_2} - \frac{1}{t_1})^2} \times (\alpha - 1)[e^{\frac{2\alpha_2}{t_1}}(\frac{1}{t} - \frac{1}{t_1})^2 + e^{\frac{2\alpha_2}{t_2}}(\frac{1}{t} - \frac{1}{t_2})^2]. \tag{2.14}$$

Plugging in t_2 in (2.14), we find $d'_{\boldsymbol{\theta}}(\xi, t_2) \geq 0$, which implies $t_2 = t_{max}$.

Next, note that ξ will be a locally D-optimal design if $d_{\boldsymbol{\theta}}(\xi, t)$ has a local maximum value at t_1. Setting $d'_{\boldsymbol{\theta}}(\xi, t) = 0$, we find $t_1 = \frac{\alpha_2 t_2}{\alpha_2 + t_2}$. If $\frac{\alpha_2 t_2}{\alpha_2 + t_2} < t_{min}$, then $d'_{\theta}(\xi, t) < 0$ for $t \in [t_{min}, t_{max})$. It follows that $d_{\theta}(\xi, t)$ is a decreasing function in $[t_{min}, t_{max}]$; thus $d_{\theta}(\xi, t)$ will have its local maximum at t_{min}. Consequently, $t_1^* = max(\frac{\alpha_2 t_2}{\alpha_2 + t_2}, t_{min})$.

Also, by the definition of D-optimal design, the above design identifies an ellipsoid which contains all points on $\mathbf{v}_1(t), t \in [t_{min}, t_{max}]$ with minimum volume. There will be two points, $\frac{\alpha_2 t_{max}}{\alpha_2 + t_{max}}, t_{max}$ at the boundary of the above ellipsoid. □

2.3 A geometric approach to D-Optimal designs and its application to models M_2 through M_4

The geometric interpretation of D-optimal designs has a long history. Silvey (1972), Sibson (1972), Silvey and Titterington (1973), Ford and Silvey (1980), Haines (1993) and Vandenberghe et al. (1998) have studied the geometric properties of D-optimal designs, which actually follow from the dual of the optimization problem given by (2.4).

Let $\boldsymbol{\theta}$ be a vector of k unknown parameters and suppose the $k \times k$ information matrix $\mathbf{I}$ can be expressed in the form $\mathbf{v}\mathbf{v}'$, where $\mathbf{v}$ is a member of a compact subset $\mathcal{V}$ of $\mathbb{R}^k$. Then, the definition of D-optimality described earlier in Section 2.2 can be extended as follows (Titterington, 1975): the design that maximizes $\log\det \int_{\mathcal{V}} \mathbf{v}\mathbf{v}'\xi(d\mathbf{v})$ is D-optimal for estimating $\boldsymbol{\theta}$. Sibson (1972) showed that the resulting information matrix defines the ellipsoid of smallest volume, centered at the origin, that contains $\mathcal{V}$. It is also known (Vandenberghe et al., 1998) that the points that lie on the boundary of the minimum volume ellipsoid are the only ones that have non-zero mass. In other words, the points on the boundary of the ellipsoid will be the only support points of the design.

Note that, all the D-optimal designs derived in this Section will actually be locally D-optimal designs, which is the optimal design if the parameters are known to be close to the true value (Chernoff, 1953); however, for convenience, we shall drop the word "locally" and simply refer to them as D-optimal designs.

2.3.1 D-optimal design for model M_2

Consider the growth function $g_{\boldsymbol{\theta}}(t)$ given by (2.3) with $\alpha_1 = 10, \alpha_2 = 1.2, t_{\min} = 5, t_{\max} = 40$ and $t_0 = 5$. The left panel of Figure 2.1 shows the plot of $Y(t)$ given by (2.1) with $\sigma = 0.1$. For this model, it follows from (2.8) that $\mathbf{v} = \mathbf{v}_2 = (v_2^1, v_2^2)$ represents a two dimensional vector in the compact set $\mathcal{V}_2 = \{(v_2^1(t), v_2^2(t)) : t_{\min} \leq t \leq t_{\max}\}$, where

$$v_2^1(t) = \begin{cases} e^{-\frac{\alpha_2}{t}}, & t < t_0 \\ e^{-\frac{\alpha_2}{t_0}}\left(1 - \frac{\alpha_2}{t_0} + \frac{\alpha_2 t}{t_0^2}\right), & t \geq t_0 \end{cases},$$

and,

$$v_2^2(t) = \begin{cases} -\frac{\alpha_1}{t} e^{-\frac{\alpha_2}{t}}, & t < t_0 \\ -\frac{\alpha_1}{t_0} e^{-\frac{\alpha_2}{t_0}}\left(2 - \frac{\alpha_2}{t_0} - \frac{t}{t_0} + \frac{\alpha_2 t}{t_0^2}\right). & t \geq t_0 \end{cases}.$$

The right panel of Figure 2.1 shows all the points in the set $\mathcal{V}_2$ defined above and the minimum volume ellipse centered at the origin that contains all these points. Clearly, the ellipse must touch only two points on the curve; one of them is $(x_1, y_1) = (v_2^1(t_{\max}), v_2^2(t_{\max}))$, which lies at one extreme of the curve. It is obvious that the second point will be of the form $(x_2, y_2) = (v_2^1(t), v_2^2(t))$ for some $t \leq t_0$. Minimizing the volume of the ellipsoid passing through (x_1, y_1) and (x_2, y_2) with respect to t, we obtain the second support point of the D-optimal design and hence arrive at the following Theorem:

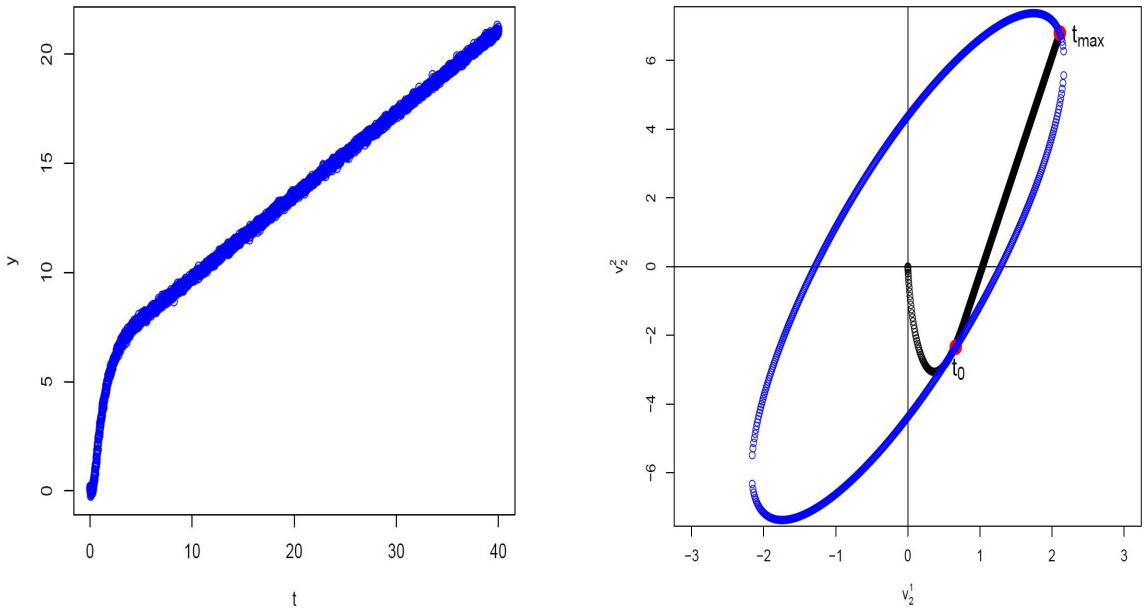

Figure 2.1: Response function (left panel) and Minimum volume ellipse that contains $\mathcal{V}$ (right panel) for Model M_2 with $\alpha_1 = 10, \alpha_2 = 1.2, \sigma = 0.1, t_{\min} = 0.1, t_{\max} = 40$ and $t_0 = 5$.

Theorem 2. *Define*

$$\tau = \frac{\alpha_2}{1 - \alpha_2 \frac{(t_{\max} - 2t_0)t_0 - \alpha_2(t_{\max} - t_0)}{t_0(t_0^2 + \alpha_2(t_{\max} - t_0))}}. \tag{2.15}$$

The D-optimal design for model M_2 is a balanced two-points design with unique support at t_1^*, t_2^*, where

$$t_2^* = t_{max} \quad and \quad t_1^* = max(\tau, t_{min})$$

Proof. Since the volume of the ellipse is given by $1/(AC - B^2)$, to find the minimum volume ellipse we need to find A, B and C that maximizes $AC - B^2$ subject to $Ax_1^2 + 2Bx_1y_1 + Cy_1^2 = 1$ and $Ax_2^2 + 2Bx_2y_2 + Cy_2^2 = 1$. Point (x_1, y_1) is $\mathbf{v}_2(t_{max})$ and (x_2, y_2) is $\mathbf{v}_2(t_1^*)$. The objective function can be written as

$$S = (AC - B^2) + L_1(Ax_1^2 + 2Bx_1y_1 + Cy_1^2 - 1) + L_2(Ax_2^2 + 2Bx_2y_2 + Cy_2^2 - 1),$$

where L_1 and L_2 are Lagrangian multipliers. Differentiating S with respect to A, B and C and equating them to zero, we obtain

$$A = -(L_1y_1^2 + L_2y_2^2); \quad B = L_1x_1y_1 + L_2x_2y_2; \quad C = -(L_1x_1^2 + L_2x_2^2). \tag{2.16}$$

Substituting the values of A, B and C in the two constraints, we obtain

$$L_1 = L_2 = -1/(x_1y_2 - x_2y_1)^2. \tag{2.17}$$

The result follows by re-substituting (2.17) into (2.16). It is easy to check that the determinant of the Hessian matrix is -2, which means the objective function is

maximized. Substituting the final expressions of A, B and C into $1/(AC - B^2)$, the volume of the ellipse is obtained as $(x_1y_2 - x_2y_1)^2$.

Here we have $x_1 = v_2^1(t_{\max}) = P \times R$ and $y_1 = v_2^2(t_{\max}) = (-\alpha_1/t_0)P \times Q$, where

$$\begin{aligned} P &= e^{-\alpha_2/t_0}, \\ R &= 1 - \frac{\alpha_2}{t_0} + \frac{\alpha_2 t_{\max}}{t_0^2}, \text{ and} \\ Q &= 2 - \frac{\alpha_2}{t_0} - \frac{t_{\max}}{t_0} + \frac{\alpha_2 t_{\max}}{t_0^2}. \end{aligned}$$

Since $t_0 < t_{\max}$, it follows that $Q < R$.

Also, we have $x_2 = e^{-\alpha_2/t_1^*}$ and $y_2 = (-\alpha_1/t_1^*)e^{-\alpha_2/t_1^*}$ where $t_1^* \leq t_0$.

In order for the ellipsoid to contain $\mathbf{v}_2(t), t \in [t_{min}, t_{max}]$, the ellipsoid should be tangent with $\mathbf{v}_2(t)$ at t_1^*, unless $t_1^* = t_{min}$. Then we have:

$$(Ax_2 + By_2) + (Bx_2 + Cy_2)\frac{dy}{dx}(x_2) = 0. \tag{2.18}$$

By substituting A, B, x_2, y_2 and $\frac{dy}{dx}(x_2) = \left(\frac{\alpha_1}{\alpha_2} - \frac{\alpha_1}{t_1^*}\right)$ into Equation (2.18), we have:

$$\left(-\frac{\alpha_1}{t_0}PQ - PR\left(\frac{\alpha_1}{\alpha_2} - \frac{\alpha_1}{t_1^*}\right)\right) = 0. \tag{2.19}$$

Equation (2.19) has only one zero point:

$$\tau = \frac{\alpha_2}{1 + \frac{\alpha_2}{t_0}\frac{Q}{R}}. \tag{2.20}$$

So if $t_{min} < \tau$, $t_1^* = \tau$, otherwise $t_1^* = t_{min}$.

It remains to prove the ellipsoid above contains all points of $\mathbf{v}_2(t), t \in [t_{min}, t_{max}]$. Consider function:

$$f(t) = Ax(t)^2 + 2Bx(t)y(t) + Cy^2(t) - 1, t \in [t_{min}, t_0] \tag{2.21}$$

The derivative of $f(t)$ is given by:

$$f'(t) = -\frac{R\frac{\alpha_2}{t} - R - \frac{Q}{t_0}\alpha_2}{t^2\left(\frac{Q}{t_0} - \frac{R}{t}\right)}. \tag{2.22}$$

The function $f'(t)$ only has one zero point, which is τ. It is easy to show that $f'(t) < 0, t > \tau$ and $f'(t) > 0, t < \tau$. Then $f(t)$ will get its maximum at $t_1^* = \tau \vee t_{min}$ for $t \in [t_{min}, t_0]$. Consequently, the above ellipsoid contains all points on $\mathbf{v}_2(t), t \in [t_{min}, t_0]$.

The $\mathbf{v}_2(t), t_0 < t \leq t_{max}$ is a straight line. The ellipsoid contains both $\mathbf{v}_2(t_0)$ and $\mathbf{v}_2(t_{max})$, so the ellipsoid will contain all the points between t_0 and t_{max}.

In conclusion, we have find the ellipsoid which contains all points, center at the original with minimum volumes. □

Remark 1. *Clearly, as* $t_{\max} \to t_0$*, model* M_2 *tends to model* M_1 *with maximum feasible time* t_0*. Note that*

$$\lim_{t_{\max} \to t_0} t_1^* = max(\frac{\alpha_2 t_0}{\alpha_2 + t_0}, t_{\min}),$$

which shows that the limiting value of t_1^* *as* $t_{\max} \to t_0$ *is the same as* t_1^* *for model* M_1*. Thus, Theorem 1 can be deduced as a special case of Theorem 2.*

2.3.2 D-optimal design for model M_3

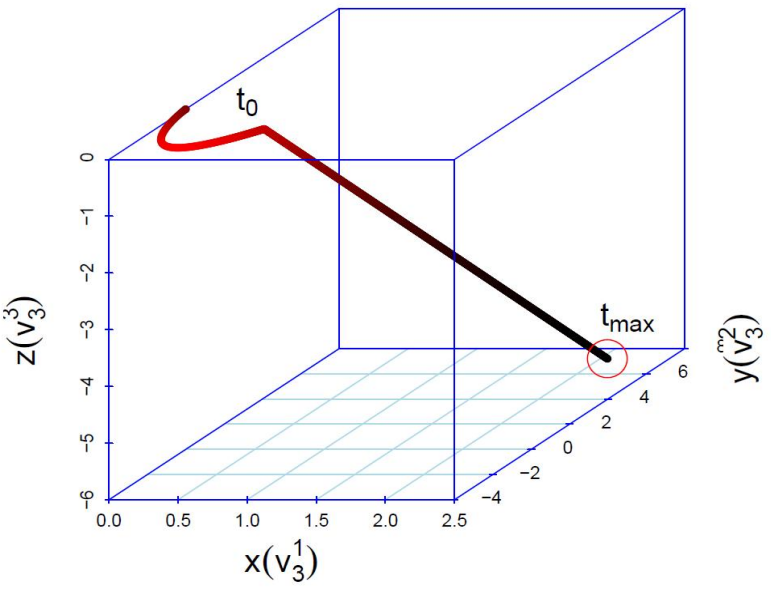

Figure 2.2: Plot of candidate points in $\mathcal{V}_3$ for Model M_3 with $\alpha_1 = 10, \alpha_2 = 1.2, \sigma = 0.1, t_{\min} = 0.1, t_{\max} = 40$.

Now we consider the case when the change-point t_0 is unknown. For this model, $\mathbf{v} = \mathbf{v}_3 = (v_3^1, v_3^2, v_3^3)$ represents a three-dimensional vector in the compact set $\mathcal{V}_3 = \{(v_3^1(t), v_3^2(t), v_3^3(t)) : t_{\min} \leq t \leq t_{\max}\}$, where $v_3^j(t), j = 1, 2, 3$ represent the components of $\mathbf{v}_3$ given by (2.9). Note that they are different for $t < t_0$ and $t \geq t_0$. Figure 2.2 shows all the points in the set $\mathcal{V}_3$ defined above. The points lie on a curve in the $X - Y$ plane for $t < t_0$, and then fall on a straight line in the 3-D space till $t = t_{\max}$. In this case, the minimum volume ellipsoid must touch three of the points in $\mathcal{V}_3$, one of which has to be $(v_3^1(t_{\max}), v_3^2(t_{\max}), v_3^3(t_{\max}))$, i.e., the point corresponding to $t = t_{\max}$. It is not difficult to see that if the minimum volume ellipsoid is made to pass through this point, then the maximization problem reduces to maximization of

the projection of the ellipsoid on the $X - Y$ plane. Consequently, we arrive at the following Theorem:

Theorem 3. *The D-optimal design for model M_3 is a three points balanced design with support at the following points:*

$$t_1^* = max(\frac{\alpha_2 t_0}{\alpha_2 + t_0}, t_{min}), \quad t_2^* = t_0, \quad t_3^* = t_{max}$$

.

Proof. In this case, we have a ellipsoid in three dimensions, which can be expressed as:

$$Ax^2 + By^2 + Cz^2 + 2Dxy + 2Exz + 2Fyz = 1. \tag{2.23}$$

Let the matrix

$$\begin{pmatrix} A & D & E \\ D & B & F \\ E & F & C \end{pmatrix}$$

be denoted by $\mathbf{S}$. Then the volume of the ellipsoid is $\det[S^{-1}]$, which means, to obtain the minimum volume ellipsoid, we need to maximize

$$\det[\mathbf{S}] = (AB - D^2)C - BE^2 + 2DEF - AF^2.$$

Now we know that the ellipsoid must pass through the point that corresponds to

$t = t_{\max}$. Denoting this point by(x_M, y_M, z_M), we have from (2.23)

$$C = \frac{1 - Ax_M^2 - By_M^2 - 2Dx_My_M - 2Ex_Mz_M - 2Fy_Mz_M}{z_M^2}. \tag{2.24}$$

Substituting the above expression of C into $\det[\mathbf{S}]$, we have

$$\begin{aligned}
\det[\mathbf{S}] &= (AB - D^2)\Big(\frac{1 - Ax_M^2 - By_M^2 - 2Dx_My_M}{z_M^2}\Big) \\
&- 2E\frac{x_M}{z_M}(AB - D^2) - 2F\frac{y_M}{z_M}(AB - D^2) - BE^2 + 2DEF - AF^2 \\
&= (AB - D^2)\Big(\frac{1 - Ax_M^2 - By_M^2 - 2Dx_My_M}{z_M^2}\Big) \\
&-2E\beta_1 - 2F\beta_2 - BE^2 + 2DEF - AF^2,
\end{aligned} \tag{2.25}$$

where

$$\beta_1 = \frac{x_M}{z_M}(AB - D^2), \text{ and } \beta_2 = \frac{y_M}{z_M}(AB - D^2).$$

Taking the derivative of $\det[\mathbf{S}]$ with respect to E, F, we have:

$$\begin{aligned}
\frac{\partial \det[\mathbf{S}]}{\partial E} &= -2EB + 2DF - 2\beta_1, \\
\frac{\partial \det[\mathbf{S}]}{\partial F} &= -2FA + 2DE - 2\beta_2.
\end{aligned}$$

Equating the above partial derivatives to zero, we have:

$$\begin{pmatrix} -B & D \\ D & -A \end{pmatrix} \begin{pmatrix} E \\ F \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \tag{2.26}$$

Substituting E and F from (2.26) into (2.25), we have:

$$\det[\mathbf{S}] = \frac{AB - D^2}{z_M^2}. \tag{2.27}$$

Since z_M is a constant, the optimization problem in three dimension reduces to maximization of $AB - D^2$, which is the projection of the ellipsoid on the $X - Y$ plane. Thus, applying the result of Theorem 1 with $t_{\max} = t_0$, the proof immediately follows. Also, since the projection of the $\mathbf{S}$ on $X - Y$ plane contains $\mathbf{v}_3(t_0)$, then the ellipsoid $\mathbf{S}$ will contain points on straight line $\mathbf{v}_3(t), t \in [t_0, t_{max}]$. In conclusion, the ellipsoid $\mathbf{S}$ contains $\mathbf{v}_3(t), t \in [t_{min}, t_{max}]$. □

Remark 2. *This result is quite intuitive. By Theorem 1, the D-optimal design for the exponential part of the curve has to be a balanced two-point design with support at points* $\{t_1^*, t_2^*\}$*, and it is obvious that for the linear part the D-optimal design is again a balanced design with support at points* $\{t_2^*, t_3^*\}$*. As in Theorem 2, if we consider the limit of this design as* $t_{\max} \to t_0$*, it is seen to converge to the D-optimal design for model* M_1*.*

2.3.3 D-optimal design for model M_4

Finally we consider the case when we drop the assumption of continuity of the first order derivative of $g_{\boldsymbol{\theta}}(t)$. For this model, $\mathbf{v} = \mathbf{v}_4 = (v_4^1, v_4^2, v_4^3, v_4^4)$ represents a four-dimensional vector in the compact set $\mathcal{V}_4 = \{(v_4^1(t), v_4^2(t), v_4^3(t), v_4^4(t)) : t_{\min} \leq t \leq t_{\max}\}$, where $v_4^j(t), j = 1, 2, 3, 4$ represent the components of $\mathbf{v}_4$ given by (2.11) and, as before, are different for $t < t_0$ and $t \geq t_0$. Similar to model M_3, the points lie on a curve in the $X-Y$ plane for $t < t_0$. For $t \geq t_0$, they fall on a straight line in the four-dimensional space till $t = t_{\max}$. In this case, the minimum volume ellipsoid must touch four of the points in $\mathcal{V}_4$, one of which has to be $(v_4^1(t_{\max}), v_4^2(t_{\max}), v_4^3(t_{\max}), v_4^4(t_{\max}))$, i.e., the point corresponding to $t = t_{\max}$. After some tedious algebra, it can be seen that if the minimum volume ellipsoid is made to pass through this point, then the maximization problem reduces to maximization of the projection of the ellipsoid on the $X - Y - Z$ plane. Consequently, we arrive at the following Theorem:

Theorem 4. *The D-optimal design for model M_4 is a four points balanced design with support at the following points:*

$$t_1^* = max(\frac{\alpha_2 t_0}{\alpha_2 + t_0}, t_{min}), \quad t_2^* = t_0^-, \quad t_3^* = t_0, \quad t_4^* = t_{max},$$

where

$$t_0^- = \sup_{t_{min} < t < t_0} \{t : g'_{\boldsymbol{\theta}}(t) = \frac{\alpha_1 \alpha_2}{t^2} e^{-\alpha_2/t}\}.$$

Proof. In this case, we have an ellipsoid in $\mathbb{R}^4$, which can be expressed as:

$$Ax^2 + By^2 + Cz^2 + Du^2 + 2Exy + 2Fxz + 2Hyz + 2Gxu + 2Iyu + 2Jzu = 1. \quad (2.28)$$

Let the matrix

$$\begin{pmatrix} A & E & F & G \\ E & B & H & I \\ F & H & C & J \\ G & I & J & D \end{pmatrix}$$

be denoted by $\mathbf{S}$. For simplicity, we also refer to the ellipsoid defined by the above matrix as ellipsoid $\mathbf{S}$. Then the volume of the ellipsoid is $\det[\mathbf{S}^{-1}]$, which means, to obtain the minimum volume ellipsoid, we need to maximize $\det[\mathbf{S}]$.

The graph of $\mathbf{v}_4(t)$ represents a straight line for $t \geq t_0$. Let (x_M, y_M, z_M, u_M) denote the point at which the ellipsoid intersects t_{max}. Then,

$$\begin{aligned} D &= \frac{1}{u_M^2} \\ &\times 1 - Ax_M^2 - By_M^2 - Cz_M^2 - 2Ex_My_M - 2Fx_Mz_M - 2Hy_Mz_M - 2Gx_Mu_M - 2Iy_Mu_M - 2Jz_Mu_M \end{aligned}$$

Now, $det[\mathbf{S}]$ can be written as:

$$\begin{aligned} det[\mathbf{S}] &= Ddet[\mathbf{S}_1] \\ &- J\{A(BJ - HI) + E(IF - EJ) + G(EH - BF)\} \\ &+ I\{A(HJ - CI) + F(IF - EJ) + G(EC - HF)\} \\ &- G\{E(HJ - CI) + F(HI - BJ) + G(BC - H^2)\} \end{aligned}$$

where

$$\mathbf{S}_1 = \begin{pmatrix} A & E & F \\ E & B & H \\ F & H & C \end{pmatrix}$$

Plugging in D, we have:

$$\begin{aligned} det[\mathbf{S}] &= \frac{1 - Ax_M^2 - By_M^2 - Cz_M^2 - 2Ex_My_M - 2Fx_Mz_M - 2Hy_Mz_M}{u_M^2} det\{\mathbf{S}_1\} \\ &- 2G\beta_1 - 2I\beta_2 - 2J\beta_3 \\ &- J\{A(BJ - HI) + E(IF - EJ) + G(EH - BF)\} \\ &+ I\{A(HJ - CI) + F(IF - EJ) + G(EC - HF)\} \\ &- G\{E(HJ - CI) + F(HI - BJ) + G(BC - H^2)\} \end{aligned} \tag{2.29}$$

Setting $\frac{\partial det[\mathbf{S}]}{\partial G} = 0$, $\frac{\partial det[\mathbf{S}]}{\partial I} = 0$, $\frac{\partial det[\mathbf{S}]}{\partial J} = 0$, we have:

$$\begin{aligned} \beta_1 &= -EHJ + BFJ + EIC - HIF - GBC + GH^2 \\ \beta_2 &= -EFJ + HJA - CIA + GEC - GHF + IF^2 \\ \beta_3 &= -ABJ + AHI - EIF - GEH + GBF + JE^2 \end{aligned} \tag{2.30}$$

Plugging in (2.30) into (2.29), we have:

$$det[\mathbf{S}] = \frac{det[\mathbf{S}_1]}{u_M^2}, \tag{2.31}$$

which implies that it suffices to maximize $det[\mathbf{S}_1]$. Note that $\boldsymbol{S}_1$ corresponds to the projection of ellipsoid $\mathbf{S}$ into the x-y-z plane. It consists of two parts, the first of which is the information vector $\mathbf{v}_4(t)$ when $t < t_0$, and the other is $\mathbf{v}_4(t_0)$. Then, proceeding as in the proof of Theorem 3, we can find ellipsoid $\mathbf{S}_1$ with minimum volume which contains $\mathbf{v}_4(t), t \in [t_{min}, t_0]$. Because of the discontinuity of information vector $\mathbf{v}_4(t)$ at t_0, there will be three points $\frac{\alpha_2 t_0}{\alpha_2 + t_0}, t_0^-$ and t_0 on the boundary of the ellipsoid $\mathbf{S}_1$, where $t_0^- = \sup_{t_{min}<t<t_0}\{t : g'_{\boldsymbol{\theta}}(t) = \frac{\alpha_1\alpha_2}{t^2}e^{-\alpha_2/t}\}$. Also, since $\mathbf{v}_4(t), t \in [t_0, t_{max}]$ is a straight line and $\mathbf{S}_1$ contains $\mathbf{v}_4(t), t \in [t_{min}, t_0]$, the resulting ellipsoid $\mathbf{S}$ contains all points on $\mathbf{v}_4(t), t \in [t_{min}, t_{max}]$. □

Remark 3. *The difference between t_0 and t_0^- is due to the discontinuity of the first order derivative of the growth curve. The locally D-optimal design for M_4 can be approximated by a design with 25% weight at each of the points t_1^* and t_{max}, and 50% weight in a small interval $(t_0 - \epsilon, t_0]$ including t_0.*

2.4 A sequential Bayesian strategy to generate the D-optimal design

The D-optimal designs derived in Section 2.3 depend on the actual values of the parameters in the model, which are unknown. The idea of sequential design for this type of problem is natural (Box and Hunter, 1965; Chernoff, 1973; Ford and Silvey, 1980; Hohmann and Jung, 1975; Silvey, 1980). Sequential strategies use currently available data to choose the next design points, and can broadly be divided into two categories: (i) sequential non-Bayesian strategies (Chaudhuri and Mykland, 1993)

where at each stage the local optimal design is computed at the current maximum likelihood estimates of the parameters, and (ii) sequential Bayesian strategies (Dror and Steinberg, 2008; Roy et al., 2009) where at each stage the optimal design is obtained by optimizing the expected D-optimality criterion over a suitable prior distribution for the parameters. We prefer to use a sequential Bayesian strategy for our problem mainly due to two reasons. First, the experimenters usually have some prior information about the parameters of interest (e.g., the change point t_0 is likely to occur early in the growth process) which can be readily incorporated into a Bayesian strategy. Second, statistical inference of the model parameters for small sample size is more straightforward in a Bayesian set up.

Let $\pi(\boldsymbol{\theta}, \sigma^2)$ denote a prior distribution for $\boldsymbol{\theta}$ and σ^2. Then, for model M_4, the sequential Bayesian procedure can be described in the following steps:

1. Assume n experiments have been conducted at points $t_1, \ldots, t_n$, generating observations $y_1, \ldots, y_n$. Compute the posterior distribution $\pi\left(\boldsymbol{\theta}, \sigma^2 | y_1, \ldots, y_n\right)$ based on these observations.

2. The $(n+1)$th design point t_{n+1} can be chosen by maximizing the function:

$$\int_{\Theta} log\left(det(\sum_{r=1}^{n} \mathbf{I}(\boldsymbol{\theta}, t_r) + \mathbf{I}(\boldsymbol{\theta}, t_{n+1}))\right) \pi\left(\boldsymbol{\theta}, \sigma^2 | y_1, \ldots, y_n\right) d\sigma^2 d\boldsymbol{\theta}. \qquad (2.32)$$

where $\mathbf{I}(\boldsymbol{\theta}, t) = \mathbf{v}_4 \mathbf{v}_4' / \sigma^2$ is the information matrix for model M_4. If $n = 0$, posterior distribution will be simplified to the prior distribution. The above criterion function has been used to generate Bayesian sequential D-optimal design in Roy et al. (2009).

3. Generate a new observation y_{n+1} and obtain updated posterior distribution based on paired observations $(t_1, y_1), \ldots, (t_{n+1}, y_{n+1})$.

4. Repeat steps 2-3 till convergence or budget constraints.

2.4.1 Prior distributions

For parameters α_1, α_2 and b, if no prior information is available, one can specify non-informative priors

$$\log(\alpha_1) \propto 1, \quad \log(\alpha_2) \propto 1, \quad \log(b) \propto 1.$$

with α_1, α_2, b independently distributed. It is easy to check that the posterior distributions under such prior are proper. In case the experimenter has some rough idea about the lower and upper bounds of these parameters, one can specify uniform or normal priors using those bounds. It is usually known that the change point t_0 occurs early on during the growth process. Therefore, we can assume the change point t_0 to have a lognormal prior distribution that is centered around $wt_{\min} + (1-w)t_{\max}$, where $1/2 < w < 1$. The constant w reflects the belief of the experimenter regarding how early the change point is likely to occur. The parameters μ_{t_0} and $\sigma^2_{t_0}$ of the distribution can be obtained by solving the equations

$$e^{\mu_{t_0}} = wt_{\min} + (1-w)t_{\max}, \tag{2.33}$$

$$(e^{\sigma^2_{t_0}} + 2)\sqrt{e^{\sigma^2_{t_0}} - 1} = \gamma_1, \tag{2.34}$$

where $\gamma_1(>0)$ represents the skewness of the distribution, for which the experimenter can supply a value. Note that in equation (2.33), the left hand side $e^{\mu_{t_0}}$, which represents the median of the lognormal distribution, can be replaced by the mean $e^{\mu_{t_0}+\sigma^2_{t_0}/2}$.

For example, in the experiment reported by Huang et al. (2011) where $t_{\min} = 0.5$ and $t_{\max} = 210$, if we choose $w = 3/4$ and $\gamma_1 = 1$, then equations (2.33) and (2.34) yield $\mu_{t_0} \approx 4$ and $\sigma^2_{t_0} \approx 0.315$. However, if the experimenter is not confident about how early the change point may occur, it may be pragmatic to use a uniform$[t_{\min}, t_{\max}]$ prior for t_0. For the dispersion parameter σ^2, which does not affect the locally D-optimal design in the absence of information, one can specify the noninformative prior:

$$log(\sigma^2) \propto 1$$

Again, it is easy to check the above prior leads to a proper posterior distribution.

2.4.2 Sampling from the posterior distribution, optimization and stopping rule

After the prior has been specified, one can draw samples from the posterior distribution:

$$\pi\left(\boldsymbol{\theta}, \sigma^2 | y_1, ..., y_n\right) \propto p(\boldsymbol{\theta}, \sigma^2) \prod_{i=1}^{n} f(y_i | \boldsymbol{\theta}, \sigma^2). \tag{2.35}$$

using a Monte Carlo Markov Chain (Liu, 2002). The posterior draw can be obtained from (2.35) by Gibbs Sampling, which is an efficient algorithm to generate a sequence of multivariate samples from the joint probability distribution.

Moreover, when the prior distribution is not conjugate, it is not easy to derive closed form expression of conditional distribution required by Gibbs sampling. In such case, Metropolis Hastings Algorithm (Liu, 2002) has been used. In this project, we have used Gibbs Sampling to simulate 200000 samples with a 100000 burnin and thin rate 100. Also, Gelman-Rubin statistics for posterior sample with different initial points are normally very close to 1, which indicate convergence of posterior sample is reached.

After obtaining k posterior samples $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)} \dots \boldsymbol{\theta}^{(k)}$ and $(\sigma^2)^{(1)}, (\sigma^2)^{(2)} \dots (\sigma^2)^{(k)}$, the criterion function given by (2.32) is approximated as:

$$\frac{1}{k}\sum_{i=1}^{k} log\left(det(\sum_{r=1}^{n}\mathbf{I}(\boldsymbol{\theta}^{(i)}, t_r) + \mathbf{I}(\boldsymbol{\theta}^{(i)}, t_{n+1}))\right).$$

The next design point is obtained by maximizing the above expression.

A stopping rule for the design is specified in terms of the relative error of parameter estimates. Let δ denote a pre-specified threshold for the maximum relative error of parameter estimates given by

$$\max_j \left|\frac{\hat{\theta}_j^{(n)} - \hat{\theta}_j^{(n-1)}}{\hat{\theta}_j^{(n-1)}}\right|,$$

where θ_j are the components of $\boldsymbol{\theta}$. The stopping time is then defined as

$$n_\delta = \min\{n : \max_j \left| \frac{\hat{\theta}_j^{(n)} - \hat{\theta}_j^{(n-1)}}{\hat{\theta}_j^{(n-1)}} \right| < \delta\}. \tag{2.36}$$

2.5 Comparison to the naive design through simulation studies

We now investigate whether the proposed Bayesian strategy is able to create a more efficient design compared to the naive experimental strategy adopted by Huang et al. (2011). As mentioned in Section 2.1, Huang et al. (2011) chose 19 time points that were more or less equally-spaced on $[\log(t_{\min}), \log(t_{\max})]$, where $t_{\min} = 0.5$ minute and $t_{\max} = 210$ minutes and replicated the basic experiment twice to generate 57 data points. Model M_4 with parameters $\hat{\alpha}_1 = 32.11$, $\hat{\alpha}_2 = 105.65$, $\hat{b} = 0.009$, $\hat{t}_0 = 86.67$ and $\hat{\sigma}^2 = 0.086$ provided a good fit to the data. In the simulation studies to follow, we assume this to be the true model and utilize it to generate values of the response $y(t_i)$ at design points $t_i, i = 1, 2, \ldots$ that our sequential design will generate. In our simulation study, we consider non-informative priors for $\alpha_1, \alpha_2, b, \sigma^2$ and different prior distributions for t_0 as mentioned in Section 4.1.

2.5.1 Simulation results

The D-optimal design for this model will be a balanced four points design with support at points 47.47, 86.67^-, 86.67 and 210 where 86.67^- is as defined in Theorem 4. We consider Bayesian sequential designs with three different sets of prior for t_0 : $t_0 \sim unif[0.5, 210], \log(t_0) \propto 1, t_0 \sim LN(4, .315)$, which shall respectively be

referred to as Prior 1, Prior 2 and Prior 3 henceforth. It is seen that the Bayesian sequential algorithm converges early (in about 20 runs) to the locally D-optimal design for the true values of the parameters. (See Figure 2.3).

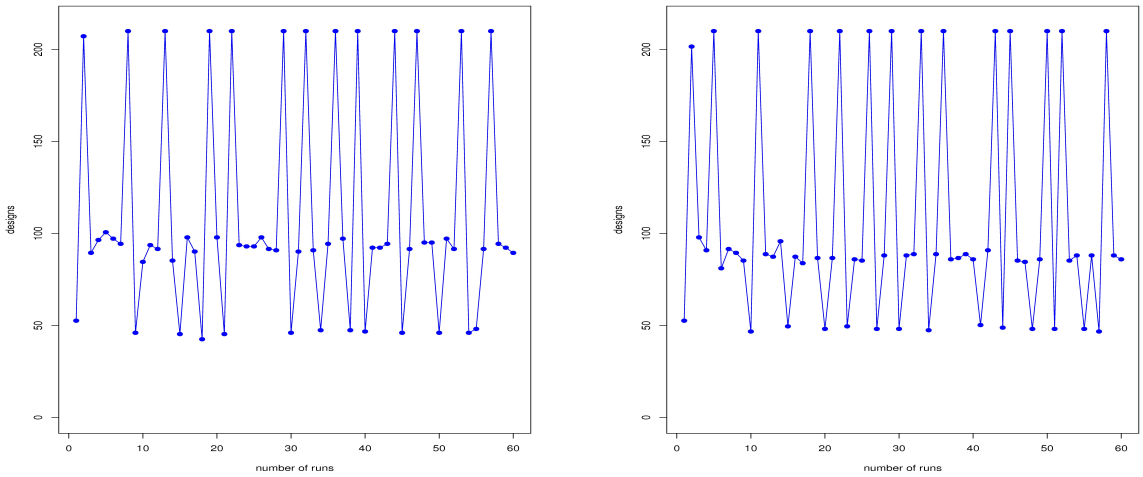

Figure 2.3: Two examples of convergence of Bayesian Sequential Design for prior 2 (left panel) prior 3 (right panel)

2.5.2 Sensitivity analysis with respect to different priors

Next we conduct simulations to assess the sensitivity of the results with respect to the choice of prior distribution. The following three performance measures are used for this analysis:

1. The rate of convergence of the design as determined by the stopping time n_δ given by (2.36) for a given threshold of relative error δ. Table 2.1 shows the median value of n_δ (obtained from 300 simulations) for $\delta = .01, .005, .001$ and three different priors.

Table 2.1: Effect of priors on median stopping time

Stopping time (n_δ)	Relative error		
	0.01	0.005	0.001
prior 1	12	19	60
prior 2	9	17	58
prior 3	6	14	53

2. For designs with a pre-fixed number of runs (n_f), the mean-squared-error (MSE) of parameter estimates over repeated simulations. Figure 2.4 shows the median MSE (obtained from 300 simulations) corresponding to all parameters for $n_f = 10, 20, 30, 40, 50, 60$ and three different priors.

3. D-efficiency measure, which, for a particular n-run design d consisting of points $\{t_1, \ldots, t_n\}$ can be defined as the ratio

$$D_{\text{eff}}(d) = \frac{\det \sum_{i=1}^{n} \mathbf{I}(\boldsymbol{\theta}^*, t_i)}{(n/4)^4 \det \sum_{t \in \{t_1^*, t_2^*, t_3^*, t_4^*\}} \mathbf{I}(\boldsymbol{\theta}^*, t)}, \tag{2.37}$$

where $\boldsymbol{\theta}^*$ denotes the true value of $\boldsymbol{\theta}$, and $t_1^*, t_2^*, t_3^*, t_4^*$ represent the support points of the locally D-optimal design (given by Theorem 4) corresponding to $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. The point t_0^- in Theorem 4 is replaced by $t_0 - \epsilon$, where ϵ is a small positive number. This measure provides a direct comparison of this Bayesian sequential design and locally D-optimal design at the true values of the parameters. The median D-efficiencies, again obtained from 300 simulations, for $n = 10, 20, 30, 40, 50, 60$ and three different priors are shown in Table 2.2.

From Figure 2.4, Tables 2.1-2.2, we find that the prior knowledge on t_0 improves the convergence, stability (particularly with respect to estimation of t_0) and D-

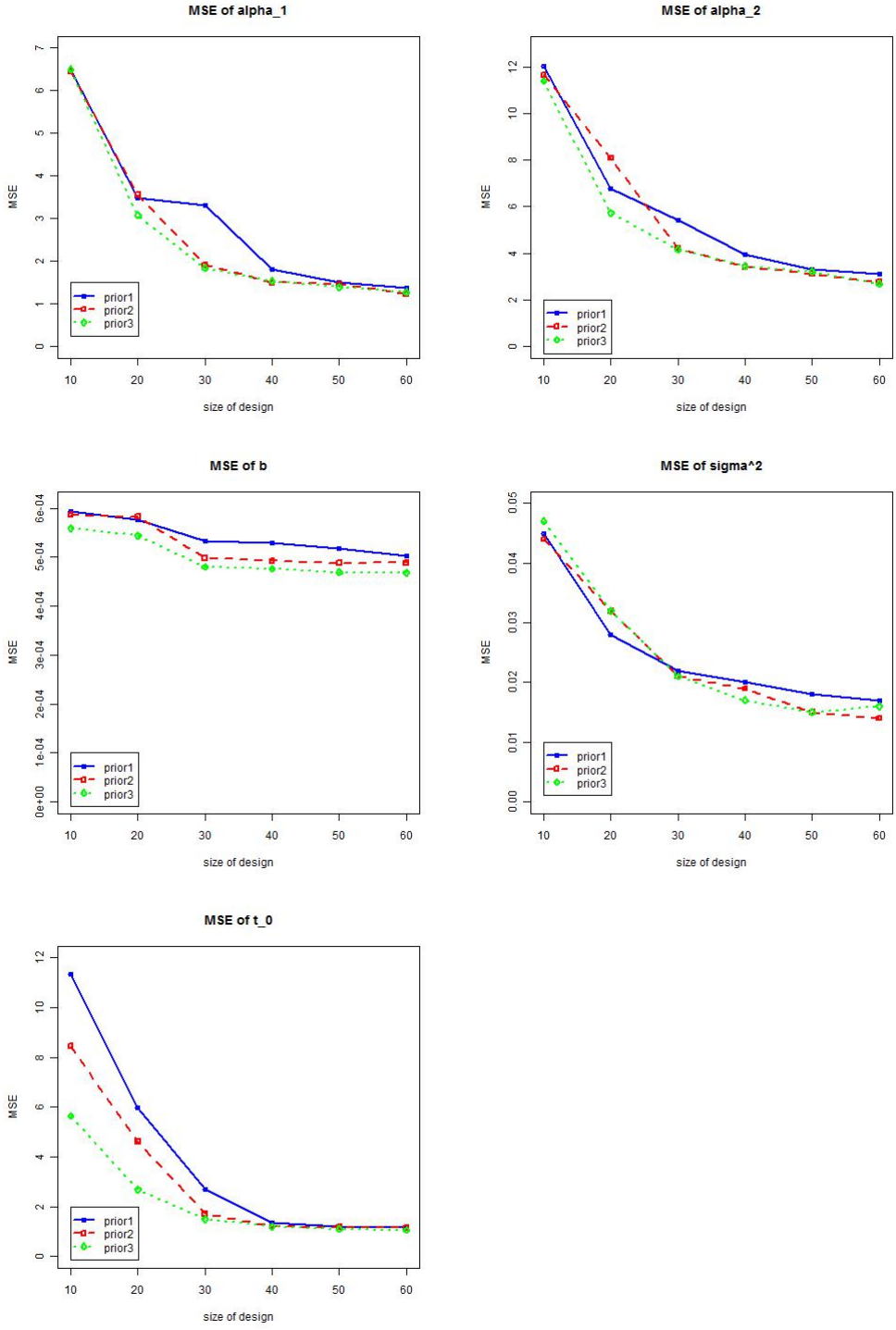

Figure 2.4: Effect of priors on median MSE of parameters for designs of fixed size n_f.

Table 2.2: Effect of priors on median D-efficiency

prior	Size of design (n)					
	10	20	30	40	50	60
prior 1	0.43	0.58	0.67	0.74	0.78	0.79
prior 2	0.45	0.60	0.68	0.75	0.78	0.80
prior 3	0.47	0.63	0.71	0.77	0.80	0.82

efficiency of the design. As expected, the MSE of estimated t_0 is smallest when a log-normal prior is used and largest when a uniform prior is used. The effect of prior distribution on estimation efficiency is more pronounced for small number of runs (≤ 20). The stopping time and D-efficiency also seem to be somewhat affected by the choice of prior distribution, although these effects are not as pronounced as the effect of prior on the MSE of t_0.

2.5.3 Comparison of the Bayesian Sequential Design and Naive Design used by Huang et al. (2011)

Finally, we compare the performance of our design with the naive design strategy adopted by Huang et al. (2011) to examine whether the proposed strategy really has the potential to meet the experimenters' requirements as stated in the introductory section.

We first compare the two designs with respect to their efficiencies of estimation of parameters for a fixed sample size. Using the same number of design points as in Huang et al. (2011) (19 for no replication, 38 and 57 for one and two replications respectively), the median root-mean-squared-error (RMSE) of each parameter, obtained from 300 simulations, is computed for both designs. The results are shown in

Table 2.3.

Table 2.3: Comparison of the two designs with respect to median RMSE of parameters

Size of design (n)	Median RMSE of parameters through repeated simulations for The naive design					The proposed design				
	α_1	α_2	t_0	b	σ^2	α_1	α_2	t_0	b	σ^2
19	4.32	9.34	5.81	9.8E-4	0.033	2.45	5.98	1.90	9.2E-4	0.035
38	3.59	7.01	3.31	9.4E-4	0.022	1.77	4.08	1.45	8.3E-4	0.025
57	2.41	4.50	1.72	8.4E-4	0.021	1.38	3.23	1.18	7.8E-4	0.020

Next we conduct simulations to compute the number of runs taken by the proposed Bayesian sequential strategy to attain the same RMSE of parameter estimates as in 19, 38 and 57-run designs adopted by Huang et al. (2011). The empirical distribution of the number of such runs obtained from 300 simulations is shown in Table 2.4.

Table 2.4: Number of runs required to attain the same efficiency as Huang et al. (2011) 19, 38 and 57-run designs

Number of design points used by Huang et al. (2011)	Quantile of number of runs for Bayesian sequential design				
	2.5%	25%	median	75%	97.5%
19	2	3	5	7	14
38	3	6	8	11	23
57	3	14	21	29	58

From Tables 2.3 and 2.4, the benefits of the proposed design over the naive strategy adopted by Huang et al. (2011) are immediately observed. The RMSEs of the estimated parameters for the proposed design are much smaller than that of the naive design. Also, the proposed design can help us achieve the same efficiency of parameter estimates in much fewer number of runs (the median number of runs for the proposed design is about one-third of the number of runs for the naive design). Thus the pro-

posed strategy is expected to be much more cost effective and efficient compared to the design adopted by Huang et al. (2011) for the fitting of M_4.

2.6 Concluding remarks

In this chapter we have derived a locally D-optimal design for the exponential-linear change point model that explains the growth of nanostructures. A Bayesian sequential strategy that converges (as verified through simulation studies) to the locally D-optimal design corresponding to the true parameter values has been proposed. Guidelines are proposed for incorporating the experimenters' perception or knowledge about the change point into a Bayesian framework through appropriate prior distributions. The effectiveness of the strategy in comparison to existing methods has been demonstrated through a simulation study.

Chapter 3

Space-filling design with unknown constraints, Introduction, Problems and Challenges

3.1 Introduction

Experiments constitute an integral component of almost every scientific endeavor. Typically, experiments are conducted by varying several input variables within a certain region, called the operable design space, and observing the response of interests. Three of the several critical challenges encountered by modern day scientists and technologists while conducting experiments are the following: (i) expensive experimentation, demanding very efficient experimentation strategies, (ii) complex response surfaces, that is, the observable output Y is a complex black-box function of input variables making mathematical modeling extremely difficult, and (iii) un-

known boundaries of a large infeasible experimental region, defined as a sub-region of the operable design space within which an experiment does not yield any meaningful result. Examples of such challenges can be found in both physical experiments as well as in computer experiments (where the response is typically deterministic). For example, Dasgupta et al. (2008) reported an example where a systematic exploration of the optimal yield conditions for Cadmium Selenide nanostructures using a 5×9 full-factorial experimental design with two input variables (temperature and pressure) was conducted. A very irregular response function and complete disappearance of morphology in several regions of the operable design space made the exploration extremely difficult.

Traditional experimental designs with multiple inputs like factorial or fractional factorial design (Wu and Hamada, 2000) are usually inappropriate for the type of complex experiments mentioned above. Regular space filling designs like Latin hypercube design or LHD (Santner et al., 2003) are appropriate for exploration of complex response surfaces, but would result in generating several design points (that do not generate any meaningful data for model fitting) in the infeasible experimental region, thereby leading to a huge waste of experimental resources. It is evident that one needs a sequential strategy robust to different models to create a set of design points that will fulfil the above objectives.

The problem of creating space-filling designs within constrained regions is relatively new in experimental design research. Stinstra et al. (2003) proposed two algorithms for computing space-filling designs for arbitrarily constrained regions where the constraints are known apriori. Henkenjohann et al. (2005) proposed a sequential

design approach for generating space-filling designs within unknown feasible regions based on the assumptions of convex boundaries, which is often too restrictive in most of the intended applications. Joseph et al. (2010) have proposed a novel sequential strategy called 'sequential minimum energy design' (SMED) which is tailor-made for sequential detection of the global optimum of a complex unknown yield function in nanostructure synthesis.

However, most of above algorithms cannot be directly used in our problem. The algorithm proposed by Stinstra et al. (2003) assumes that the boundary between the feasible and infeasible region is known. Henkenjohann et al. (2005) assume the unknown boundary is convex, which is an overly restrictive assumption in most applications. SMED, although somewhat tailor-made to carve out the infeasible region, is a global optimization algorithm. The objective of the current research is different from global optimization, in the sense that rather than an early convergence to a global optimum, a more detailed exploration of the entire feasible region is needed.

The rest of this Chapter is organized as follows: the definition of the feasible region and some intuitive examples of the problem are given in Section 3.2. An outline and general framework of the sequential designs proposed in Chapters 4 and 5 is given in Section 3.3.

3.2 Definition of a 'feasible region' and some examples

Let $y = f(\mathbf{x}) \geq 0$ denote a continuous response function of input variables $\mathbf{x} = (x_1, ..., x_p), 0 \leq x_i \leq 1, i.e., x_i \in [0,1], i = 1, \cdots, p$. The design space Ω is a $p-$dimensional unit hypercube $[0,1]^p$. Here we assume that the design space is finite and countable. The infeasible region, which is the region with no meaningful result, can be determined by a threshold value c. We use the notation $D = \{\mathbf{x} : f(\mathbf{x}) > c\}$ to denote the feasible region.

An example of the feasible region defined above can be found in Joseph et al. (2010), and is shown in Figure 3.1. This figure shows the yield of nano structures on a scale of 0 and 1 for different temperature and pressure conditions. The temperature and pressure variables are both scaled to $[0,1]$. We observe a very irregular response function and zero yield in several regions of the operable design space. Such zero-yield regions are infeasible regions, which experimenters are keen to avoid. Thus, in this example, the threshold $c = 0$. There are large infeasible region (deep green region). Moreover, the shape of feasible region is non-convex and there are small infeasible regions embedded within feasible regions.

We now cite an example for computer experiments where the output is deterministic. Considering the problem of sampling from computationally expensive density functions. An example of such probability density function is non-normal bivariate

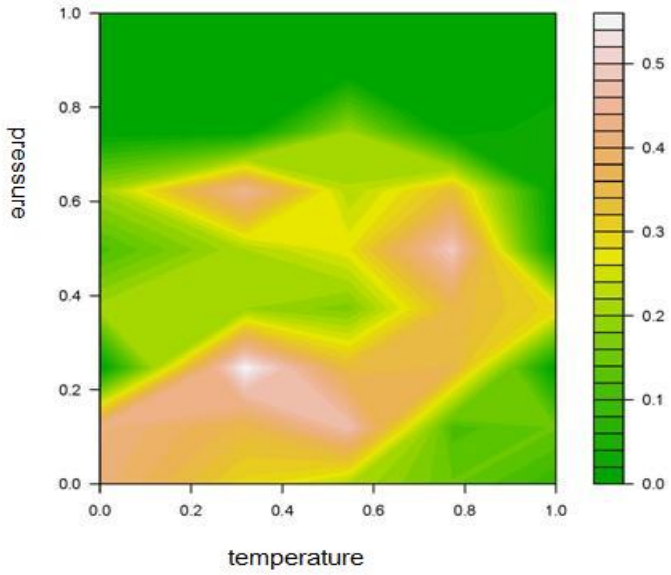

Figure 3.1: Infeasible region (deep green) for nanostructure growth.

density that is conditionally normal from Gelman and Meng (1991), given by:

$$f(x_1, x_2) \propto exp\left[-\frac{1}{2}\left(x_1^2 x_2^2 + x_1^2 + x_2^2 - 8x_1 - 8x_2\right)\right], x_1, x_2 \in \mathbb{R} \tag{3.1}$$

The contour plot of the above function is shown in Figure 3.2.

Assume that the closed form expression of the true density is unknown, however, it can be evaluated at any value of x_1, x_2. Further assume that the design space $\Omega = [-2, 6] \times [-2, 6]$. We observe that there are large infeasible regions with extremely low probability density (deep blue part) which should be avoided while sampling. Then the threshold c will be determined by the function values, 0.005 for example. We also observe that the shape of feasible region is complicated and nonconvex.

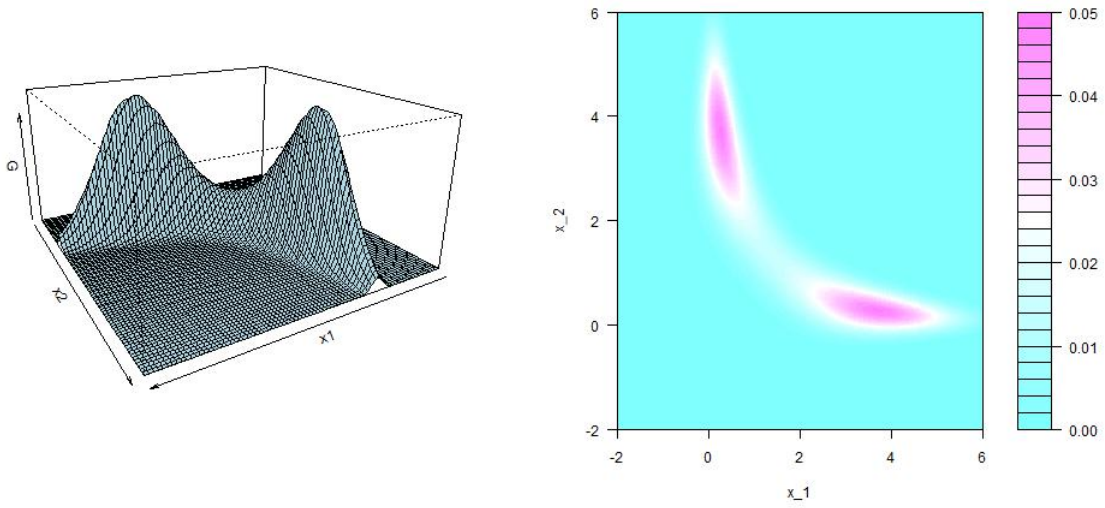

Figure 3.2: Contour plot of complex probability density function, Gelman and Meng (1991).

The above two examples give intuitive descriptions of complex input output relationships and high irregularity of the infeasible region. In the rest of this thesis, we make two simplifying assumptions on the design space: (i) The response is nonnegative, (ii) The response is deterministic. We explore design strategies theoretically and examine their effectiveness under the stated assumptions.

3.3 An outline of the proposed sequential strategies

Due to the highly complex input-output relationships and irregularity of the feasible region, a sequential strategy which can adaptively update information on the response surface is a natural choice in conducting experiment design. As in most

sequential strategies in the experimental design literature, our proposed strategies require an initial design with n_0 points. Subsequently, the design algorithm selects one point at a time by optimizing a certain design criterion. Each iteration of the algorithm has the following two stages:

1. Function evaluation: The value of the response function at every single unsampled candidate point is evaluated. The predicted value is used to distinguish between the feasible and infeasible regions.

2. Selection of the next point: The next point is selected on the basis of a pre-defined criterion function.

In this thesis, two different sequential strategies will be used. In the first strategy, an inverse distance weighting interpolator (Joseph and Kang, 2011) is used in function evaluation and a maximin space-filling criterion (Johnson et al., 1990) is used to select the next design point. In the second strategy, the response surface is modeled, assuming it to be a sample path of a Gaussian process, and the Kriging method is used for function evaluation. Two criteria related to prediction errors, *Integrated Mean Square Errors* (IMSE) and *Maximum Mean Square Errors* (MMSE) are used to select the next design point. The two strategies will be referred to as IDW-MM and GP-IMSE/MMSE henceforth and are described in detail in Chapters 4 and 5.

Chapter 4

The IDW-MM algorithm for deterministic functions

4.1 Introduction

In this Chapter, we propose a simple and intuitive algorithm to generate a space-filling design within the unknown feasible region and examine its properties through analytical derivations and simulations. As mentioned in Section 3.3, the algorithm involves two stages – (i) Functional evaluation at unsampled points, and (ii) Optimization of a criterion to select a new design point at each iteration.

Because we assume that the response is deterministic, function evaluation involves finding an interpolating function through the data points. Different interpolating functions (Pozrikidis, 1998) have been proposed in literature. We focus on a simple interpolating method – the inverse distance weighting (IDW) – proposed by Shepard (1968). Although not as accurate as some of the more sophisticated interpolators

like Kriging (Sacks et al. 1989), it is computationally simple and robust to different models. Moreover, recently there has been renewed interested in IDW interpolators (Joseph and Kang, 2011). In this Chapter, an adjustment of IDW interpolator, which only uses its near neighborhood, will be proposed and discussed.

After the response surface is estimated, a specific criterion function should be used for selecting the next design point. Different criterion-based space-filling designs have been proposed and considered. Integrated Mean Square Error (IMSE) and Maximum Mean Square Error (MMSE) criterion-based space-filling designs on a Gaussian process model have been proposed and used in Sacks et al. (1989). Expected posterior entropy, which measures the "amount of information" in an experiment (Sacks et al., 1989), has also been used as a criterion to generate space-filling design. Designs based on the Kullback-Leibler information criterion have also been considered in Jourdana and Francob (2006). However, if a proper model is not available, model-free space-filling criteria would typically be preferred. In this Chapter, a popular space-filling criterion, maximin (MM) criterion, will be used to select the next point.

The rest of this Chapter is organized as follows: in Section 4.2, the k-nearest neighbor inverse distance weighting (IDW) interpolator is introduced and discussed. We then we briefly introduce maximin space-filling design in Section 4.3. Next, the sequential IDW-MM strategy is described in Section 4.4. In Section 4.5, some convergence properties of the sequential strategy are established. In Section 4.7, extensive simulations with standard test functions are used to evaluate the performance of the space-filling strategy.

4.2 The k-nearest neighbor inverse distance weighting interpolator: Function evaluation method for step 1

Let $y_1, \cdots, y_n$ denote the observed values of response at design points $\mathbf{x}_1, \cdots, \mathbf{x}_n$. The IDW interpolator at $\mathbf{x}$ is defined as (Joseph and Kang, 2011):

$$\hat{f}(\mathbf{x}) = \sum_{m=1}^{n} v_m(\mathbf{x}) f(\mathbf{x}_m), \tag{4.1}$$

where

$$v_m(\mathbf{x}) = \frac{w_m(\mathbf{x})}{\sum_{i=1}^{n} w_i(\mathbf{x})}, \tag{4.2}$$

for $\mathbf{x} \notin \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ and $v_m(\mathbf{x}_i) = 1$ if $i = m$ and 0 otherwise. Then by definition, $\hat{f}(\mathbf{x})$ is an interpolating function. A common choice of the weighting function is:

$$w_i(\mathbf{x}) = \frac{1}{d^2(\mathbf{x}, \mathbf{x}_i)} \tag{4.3}$$

where $d(\mathbf{x}, \mathbf{x}_i) = \{\sum_{j=1}^{p} (x_j - x_{i,j})^2\}^{1/2}$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_i$.

The major problem of IDW is its poor prediction power. As discussed in (Joseph and Kang, 2011), different modifications to IDW have been proposed. One modification is to replace the response values by locally fitted polynomials (Franke, 1989). Improvements can also be obtained by replacing the local polynomials with radial basis functions (Lazzaro and Montefusco, 2002). Further modifications discussed in Joseph and Kang (2011) includes using a probability metric instead of Euclidean dis-

tance (Lukaszyk, 2004). In Joseph and Kang (2011), regression terms have been used with IDW interpolator to achieve more precise result. In this Chapter, we propose a modification of the IDW method which is motivated by our objective of carving out the infeasible region.

Most interpolation methods perform well when the sampled points are uniformly distributed. In such cases, the distribution of points between the feasible region and the infeasible region will be quite different as the points are generated, and consequently the IDW method will lead to biased predictions. For example, when a point in the infeasible region is to be estimated, a large number of points in the feasible region will have an overwhelming impact on the estimation. Such a problem can be avoided by including only those sampled points in the IDW prediction which lie in a neighborhood of $\mathbf{x}$. Two possible modifications can be used to address this issue. Joseph and Kang (2011) considered exponential decaying weights by which the points in a neighborhood of $\mathbf{x}$ can have more influence on interpolated value $\hat{f}(\mathbf{x})$ than IDW. Here we incorporate k-nearest neighbor (k-NN) method into IDW interpolation.

The k-NN method has been widely used in classification. It is an algorithm in which an object is classified by a majority of its neighbors (Bremner et al., 2005). The k-NN method has also been used in regression analysis (Yao and Ruzzo, 2006).

We use k-nearest neighbor inverse distance weighting interpolator (k-NN IDW) for estimation of $f(\mathbf{x})$. Only k nearest samples to $\mathbf{x}$ are considered when the value of $f(\mathbf{x})$ is to be determined. The k-NN neighborhood of $\mathbf{x}$ is defined as the open ball which contains exactly k nearest neighbors $\{\mathbf{x}^{(m)}, m = 1, 2, \cdots, k\}$ and has radius

$r = max_m\{d\left(\mathbf{x}, \mathbf{x}^{(m)}\right)\}$. Let $\mathbf{x}_1, ..., \mathbf{x}_n$ denote sampled points and $y_1, .., y_n$ be the observed response value. Then the k-NN IDW interpolator at $\mathbf{x}$ is defined by (4.2), where $v_m(\mathbf{x})$ is defined as:

$$v_m(\mathbf{x}) = \frac{I_k\left(\mathbf{x}_m, \mathbf{x}\right) w_m(\mathbf{x})}{\sum_{i=1}^{n} I_k\left(\mathbf{x}_i, \mathbf{x}\right) w_i(\mathbf{x})} \tag{4.4}$$

for $\mathbf{x} \notin \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ and $v_m(\mathbf{x}_i) = 1$ if $i = m$ and 0 otherwise. The weighting function $w_m(\mathbf{x})$ is the same as (4.3) and $I_k\left(\mathbf{x}_i, \mathbf{x}\right)$ is an indicator function which takes value 1, if $\mathbf{x}_i$ falls into the k-NN neighborhood of $\mathbf{x}$, and 0 otherwise. When $f(\mathbf{x}_i)$ is to be interpolated, $v_i(\mathbf{x}_i)$ is the only nonzero weights among $v_m(\mathbf{x}_i), m = 1, 2, \cdots, n$. Then $\hat{f}(\mathbf{x}_i) = \sum_{m=1}^{n} v_m(\mathbf{x}_i) f(\mathbf{x}_m) = f(\mathbf{x}_i)$. Thus k-NN IDW is also an interpolation method.

A special case of k-NN IDW is nearest neighbor interpolation (Keys 1981), which interpolates one point by the value of its nearest neighbor. The advantage of using k-NN IDW is it will only estimate the function by its neighborhood. The sampled points far away from the point to be estimated will not be considered in the estimation. The k-NN IDW will give similar results with original IDW when the sampled points are uniformly distributed within Ω.

Another reason for using k-NN estimation is that the design process, precision estimation of the response surface is not required. Instead, it will be more important to *classify* the feasible region and infeasible region. So even though k-NN estimator loses information while it uses only part of the data set, it can successfully make the *classification*.

4.3 Selection of next design point: Criterion for step 2

Maximin and minimax distance design were proposed and discussed in Johnson et al. (1990). Moreover, Johnson et al. (1990) explored several connections between certain statistical and geometric properties of space-filling designs. Johnson et al. (1990) showed maximin distance designs are asymptotically D-optimal for Gaussian process models when the correlation between sites decreases as the distance increases. Morris and Mitchell (1995) examined maximin distance designs constructed within the class of Latin hypercube arrangements in computer experiment. Some other extensions and discussions have also been given in Mitchell et al. (1995). A maximin design tries to maximize the distance between any two points within the design set. The intuition of maximin design is, the distance between any two points in the design set should not be very small. Thus the design can maximize the sparsity of points and thus achieve space-filling property. Because of simplicity of interpretation and promising properties of maximin space-filling design, we will use this criterion in this Chapter.

According to Johnson et al. (1990), given a design space Ω, an n-points design $S = (\mathbf{x}_1, ..., \mathbf{x}_n)$ is called maximin space-filling design if:

$$S_{maximin} = argmax_S\{min_{i,j} d(\mathbf{x}_i, \mathbf{x}_j)\}. \tag{4.5}$$

where $d(\bullet, \bullet)$ is Euclidean distance.

Two simple examples of maximin design on $[0, 1]$ and $[0, 1] \times [0, 1]$ from Johnson

et al. (1990) can be found in Figure 4.1.

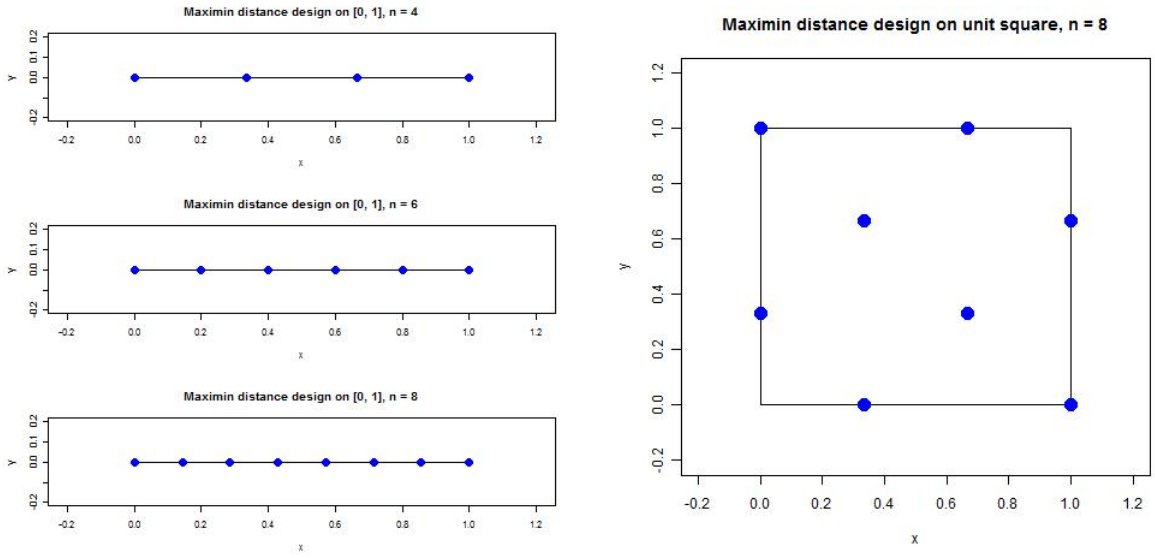

Figure 4.1: Examples of maximin distance design on $[0,1]$ and $[0,1] \times [0,1]$.

Different methods have been used to identify maximin design (Johnson et al., 1990; Morris and Mitchell, 1995). Since the boundary between feasible and infeasible region is unknown, such algorithms cannot be directly used to generate a maximin design within feasible region. We propose a greedy algorithm to sequentially approximate the maximin design in Ω. When n points $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ have already been sampled, the next point will be selected as:

$$\mathbf{x}_{n+1} = argmax_{\mathbf{x}} \min_{i=1,...,n} d(\mathbf{x}, \mathbf{x}_i). \tag{4.6}$$

The above algorithm is an approximation to the true maximin space-filling design. It is simple and requires much less computational power.

In order to evaluate the space-filling property of the sequential strategy (4.6),

we compare designs generated by this strategy and Latin hypercube designs on unit square $[0,1] \times [0,1]$ via repeated simulations. In each simulation, space-filling designs by (4.6) and Latin hypercube designs with different design sizes n ranging from 15 to 100 are generated. Let $\{\mathbf{x}_i^{SP}\}_{i=1}^n$ and $\{\mathbf{x}_i^{LHD}\}_{i=1}^n$ denote the points generated by two strategies. Then following two criteria are evaluated:

$$\phi_1(\{\mathbf{x}_i\}_{i=1}^n) = \min_{1\le i<j\le n} d(\mathbf{x}_i, \mathbf{x}_j) \tag{4.7}$$

$$\phi_2(\{\mathbf{x}_i\}_{i=1}^n) = \left(\frac{2}{n(n-1)} \sum_{1\le i<j\le n} \frac{1}{d(\mathbf{x}_i, \mathbf{x}_j)^2}\right)^{\frac{1}{2}} \tag{4.8}$$

The first performance measure $\phi_1(\{\mathbf{x}_i\}_{i=1}^n)$ evaluates minimum distance between the generated points. The second performance measure $\phi_2(\{\mathbf{x}_i\}_{i=1}^n)$ was proposed by Ba and Joseph (2011). We want $\phi_1(\{\mathbf{x}_i\}_{i=1}^n)$ to be large and $\phi_1(\{\mathbf{x}_i\}_{i=1}^n)$ to be small. The median of the two performance measures obtained from repeated simulations with respect to different n is given in Figure 4.2.

Satisfactory space-filling properties are shown by the design generated by the proposed sequential strategy. For both performance measures, the sequential strategy outperforms the Latin hypercube design. Thus, although strategy (4.6) is only an approximation to the true maximin design, it properly maintains space-filling properties.

4.4 The proposed IDW-MM algorithm

Now we can construct the sequential strategy by combining step 1 and 2 described in Section 4.2 and 4.3. First an initial design will be chosen to explore the design space

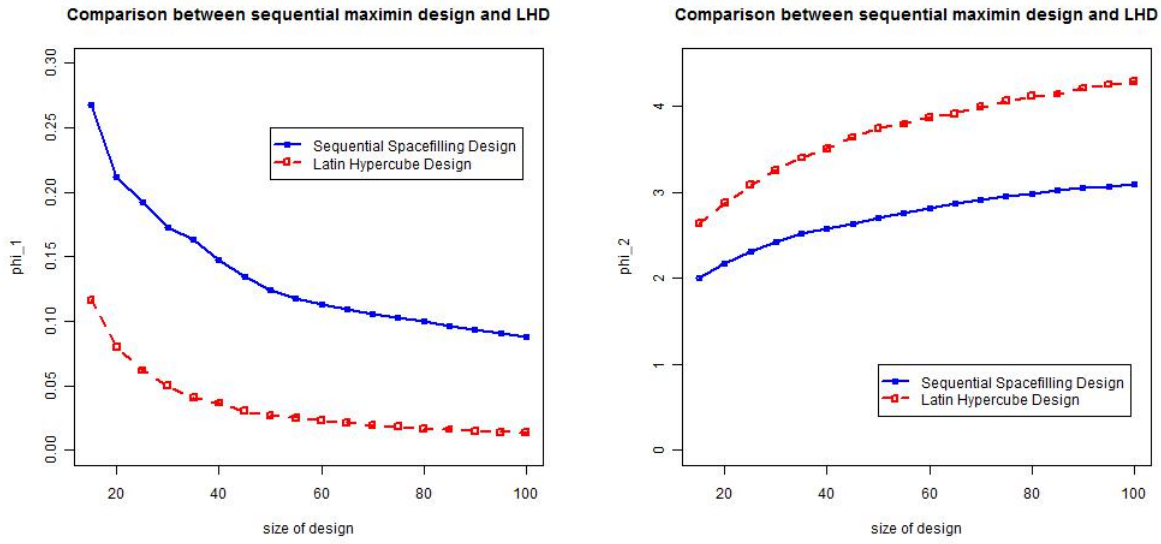

Figure 4.2: Comparison between space-filling strategy (4.6) and Latin hypercube design.

Ω. Then in each iteration of the sequential strategy, we first predict the response at all the unsampled points using the k-NN IDW interpolation method. Next we use the sequential maximin method given by (4.6) to select the next point. However, since one goal is to carve out the infeasible region, the next design point, will only be chosen from the set of points with estimated value greater than or equal to the threshold, i.e. in estimated feasible region. In this way, the selection can maintain space-filling property as well as carve out the infeasible region. The sequential algorithm to generate n^* points can be described as follows:

Step 1. Choose an initial design consisting n_0 points, using a standard space-filling strategy, e.g. Latin hypercube sampling.

Step 2. For all unsampled points $\mathbf{x}$, use the k-NN estimator to predict $f(\mathbf{x})$. Let $\hat{f}_n(\mathbf{x})$ denote the predictor of point $\mathbf{x}$ when n points have been selected.

Step 3. Select a new design point $\mathbf{x}_{n+1} = argmax_{\hat{f}_n(\mathbf{x}) \geq c} min_{i=1,\ldots,n} d(\mathbf{x}_i, \mathbf{x})$.

Step 4. Stop if $n+1 = n^*$. Else, return to Step 2.

The flow chart of the strategy can be found in Figure 4.3.

The strategy will be referred to as inverse distance weighting maximin design (IDW-MM) hereforth.

4.5 Convergence Properties of Proposed Algorithm

In this section, we examine the convergence property of the proposed sequential strategy. In the following theorems, let $\{\mathbf{x}_i\}_{i=1}^{\infty}$ denotes the sequence of points generated by the iterative IDW-MM procedure described in Section 4.4. The first n_0 points in $\{\mathbf{x}\}_{i=1}^{\infty}$ are initial design points. Also, we introduce the following notations:

- $D = \{\mathbf{x} : f(\mathbf{x}) > c\}$, the feasible region.
- $\partial D = \{\mathbf{x} : f(\mathbf{x}) = c\}$, the boundary of feasible region.
- $D^* = D \cup \partial D$.
- $\overline{D^*} = \{\mathbf{x} : f(\mathbf{x}) < c\}$, the infeasible region.
- $O_\delta(\mathbf{x}_0) = \{\mathbf{x} : d(\mathbf{x}_0, \mathbf{x}) < \delta\}$, an open ball with radius δ containing the point $\mathbf{x}_0$.
- $B(D, \delta) = \bigcup_{\mathbf{x} \in D} O_\delta(\mathbf{x})$.

Moreover, we assume the response surface is nonnegative in this section, i.e. $f(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \Omega$.

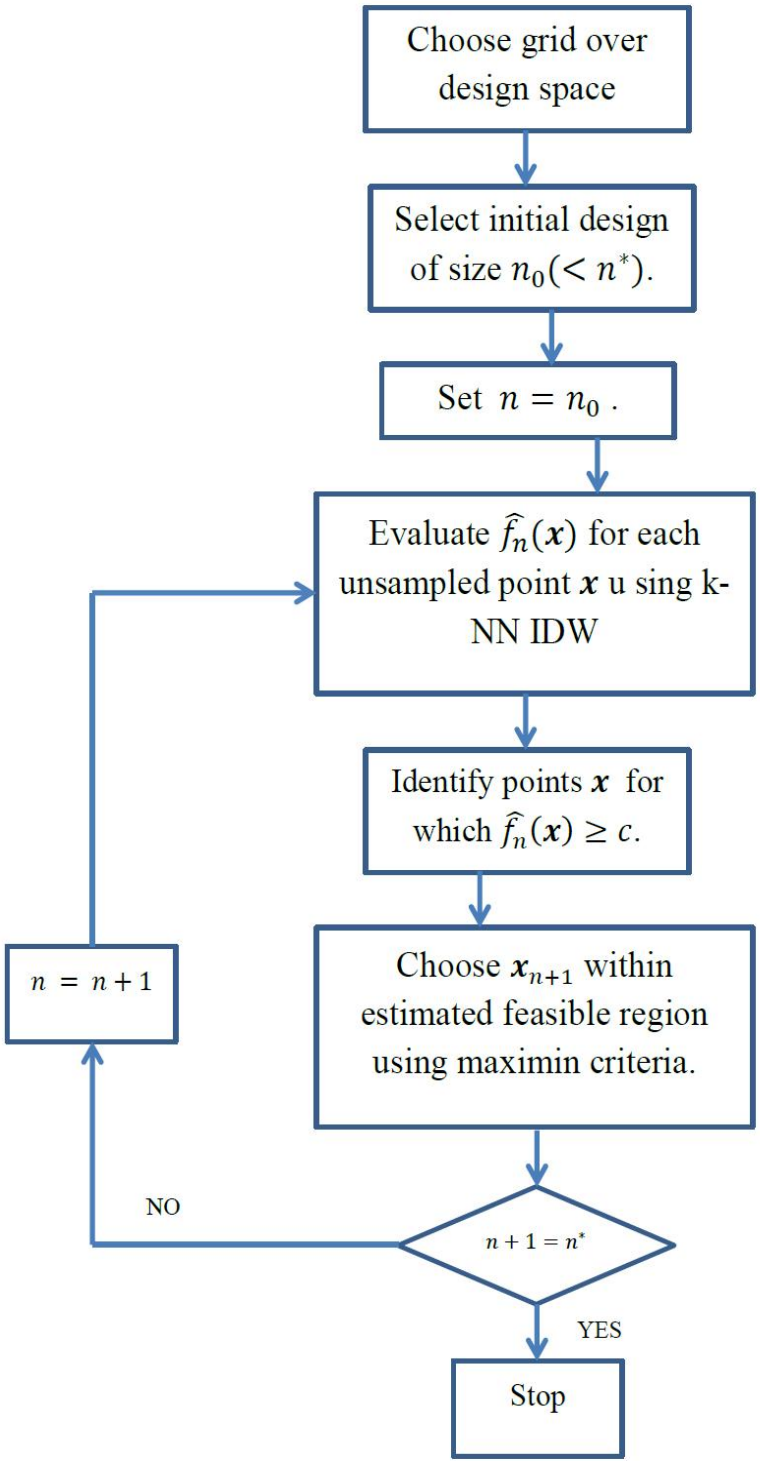

Figure 4.3: Flowchart of overall strategy.

Theorem 5. *Any arbitrary point* $\mathbf{x} \in \overline{D^*}$ *is not a limit point of the sequence* $\{\mathbf{x}_i\}_{i=1}^{\infty}$.

Proof. If $\mathbf{x}$ is a limit point, then there exists a subsequence $\{\mathbf{x}_{n_l}\}_{l=1}^{\infty} \subset \overline{D^*}$ such that $\mathbf{x}_{n_l} \to \mathbf{x}$. WLOG, we can assume the subsequence $\{\mathbf{x}_{n_l}\}_{l=1}^{\infty}$ is contained in the open ball $O_{4\delta}(\mathbf{x}) \subset \overline{D^*}$. Let N_0 be the index of the subsequence such that at least k points from $\{\mathbf{x}_{n_l}\}_{l=1}^{\infty}$ are contained in $O_{\delta}(\mathbf{x})$. When $n_l > N_0$, predicted response at all the points inside $O_{\delta}(\mathbf{x})$ is smaller than c (k-NN IDW), which is a contradiction that any of these points can be chosen. □

Remark 4. *Theorem 5 can be interpreted as follows: every point in the infeasible region is an isolated point of the sequence* $\{\mathbf{x}_i\}_{i=1}^{\infty}$.

Theorem 6. *If the feasible region* D *is path connected and there exists an initial design point* $\mathbf{x}_I$ *such that* $\mathbf{x}_I \in D$, *then* $\forall \mathbf{x} \in D$, $\mathbf{x}$ *is a limit point of the sequence* $\{\mathbf{x}_i\}_{i=1}^{\infty}$.

Proof. The proof of this theorem is relatively complicated, we prove it in several steps:

Lemma 1. *For any arbitrary point* $\mathbf{x} \in D$, *if there exists an integer* N *such that* $\hat{f}_n(\mathbf{x}) > c$ *for all* $n > N$, *then* $\mathbf{x}$ *is a limit point of the design sequence* $\{\mathbf{x}_i\}_{i=1}^{\infty}$.

Proof. Suppose $\mathbf{x}$ satisfies the condition stated in the lemma. Assume that $\mathbf{x}$ is not a limit point of $\{\mathbf{x}_i\}_{i=1}^{\infty}$. Then we can find an open ball $O_{\delta}(\mathbf{x})$ of radius δ around $\mathbf{x}$ such that no element of $\{\mathbf{x}_i\}_{i=1}^{\infty}$ lies in $O_{\delta}(\mathbf{x})$. Then, $\min_{i=1,\cdots,n} d(\mathbf{x}, \mathbf{x}_i) \geq \delta$ for all n. However, since the whole design space Ω is compact, there exists a limit point $\mathbf{x}_{\infty}$ such that a subsequence $\{\mathbf{x}_{n_l}\}_{l=1}^{\infty}$ converges to $\mathbf{x}_{\infty}$. We also have:

$$\min_{i=1,\cdots,n_l} d(\mathbf{x}_i, \mathbf{x}_{n_l}) \leq d\left(\mathbf{x}_{n_{l-1}}, \mathbf{x}_{n_l}\right) \tag{4.9}$$

Since $d\left(\mathbf{x}_{n_{l-1}}, \mathbf{x}_{n_l}\right)$ converges to 0, there exists N such that for all $n_l > N_0$:

$$\min_{i=1,\cdots,n_l} d\left(\mathbf{x}_i, \mathbf{x}_{n_l}\right) \leq \frac{\delta}{2}$$

So when $n_l > N$, $\mathbf{x}$ is in estimated feasible region while the $\min_{i=1,\cdots,n_l} d\left(\mathbf{x}, \mathbf{x}_i\right) > \min_{i=1,\cdots,n_l} d\left(\mathbf{x}_{n_l}, \mathbf{x}_i\right)$, which is contradiction that $\mathbf{x}_{n_l}$ can be chosen. □

Lemma 2. *Let $\mathbf{x}_I$ be a point chosen among the n_0 initial design points. If there exists an open ball $O_{4\delta}(\mathbf{x}_I) \subset D$ of radius 4δ among $\mathbf{x}_I$, then, there exists a point $\mathbf{x}_*$ within an open ball $O_\delta(\mathbf{x}_I)$ of radius δ among $\mathbf{x}_I$, which is a limit point of the sequence $\{\mathbf{x}_i\}_{i=1}^{\infty}$.*

Proof. Assume that no points in $O_\delta(\mathbf{x}_I)$ is a limit point. Then, the open ball $O_\delta(\mathbf{x}_I)$ can only contain a finite number of points in $\{\mathbf{x}_i\}_{i=1}^{\infty}$. Let $N_0 = \max\{i : \mathbf{x}_i \in O_\delta\left(\mathbf{x}_I\right)\}$. Then we can find $\mathbf{x}_* \in O_{\frac{(f(\mathbf{x}_I)-c)\delta}{ck}}(\mathbf{x}_I)$, such that after N_0, k-NN neighborhood of $\mathbf{x}_*$ includes $\mathbf{x}_I$. Then for $n > N_0$,

$$\hat{f}_n(\mathbf{x}_*) = \frac{\frac{f(\mathbf{x}_I)}{d(\mathbf{x}_*,\mathbf{x}_I)} + \sum_{i=1:\mathbf{x}_i\in D^*,\mathbf{x}_I\neq\mathbf{x}_i}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)f(\mathbf{x}_i)}{d(\mathbf{x}_*,\mathbf{x}_i)} + \sum_{i=1:\mathbf{x}_i\in\overline{D^*}}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)f(\mathbf{x}_i)}{d(\mathbf{x}_*,\mathbf{x}_i)}}{\frac{1}{d(\mathbf{x}_*,\mathbf{x}_I)} + \sum_{i=1:\mathbf{x}_i\in D^*,\mathbf{x}_I\neq\mathbf{x}_i}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)}{d(\mathbf{x}_*,\mathbf{x}_i)} + \sum_{i=1:\mathbf{x}_i\in\overline{D^*}}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)}{d(\mathbf{x}_*,\mathbf{x}_i)}}$$

Since $f(\mathbf{x}) \geq 0, \mathbf{x} \in \Omega$, then we have:

$$\begin{aligned}
\hat{f}_n(\mathbf{x}_*) &\geq \frac{\frac{f(\mathbf{x}_I)}{d(\mathbf{x}_*,\mathbf{x}_I)} + \sum_{i=1:\mathbf{x}_i\in D^*,\mathbf{x}_I\neq\mathbf{x}_i}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)f(\mathbf{x}_i)}{d(\mathbf{x}_*,\mathbf{x}_i)}}{\frac{1}{d(\mathbf{x}_*,\mathbf{x}_I)} + \sum_{i=1:\mathbf{x}_i\in D^*,\mathbf{x}_I\neq\mathbf{x}_i}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)}{d(\mathbf{x}_*,\mathbf{x}_i)} + \sum_{i=1:\mathbf{x}_i\in\overline{D^*}}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)}{d(\mathbf{x}_*,\mathbf{x}_i)}} \\
&\geq \frac{\frac{f(\mathbf{x}_I)}{d(\mathbf{x}_*,\mathbf{x}_I)} + \sum_{i=1:\mathbf{x}_i\in D^*,\mathbf{x}_I\neq\mathbf{x}_i}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)c}{d(\mathbf{x}_*,\mathbf{x}_i)}}{\frac{1}{d(\mathbf{x}_*,\mathbf{x}_I)} + \sum_{i=1:\mathbf{x}_i\in D^*,\mathbf{x}_I\neq\mathbf{x}_i}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)}{d(\mathbf{x}_*,\mathbf{x}_i)} + \sum_{i=1:\mathbf{x}_i\in\overline{D^*}}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)}{d(\mathbf{x}_*,\mathbf{x}_i)}}
\end{aligned}$$

Since $d(\mathbf{x}_*, \mathbf{x}_I) < \frac{(f(\mathbf{x}_I)-c)\delta}{ck}$, we have:

$$\frac{f(\mathbf{x}_I)}{d(\mathbf{x}_*, \mathbf{x}_I)} > \frac{c}{d(\mathbf{x}_*, \mathbf{x}_I)} + \frac{ck}{\delta}$$

Then:

$$\frac{\frac{f(\mathbf{x}_I)}{d(\mathbf{x}_*,\mathbf{x}_I)} + \sum_{i=1:\mathbf{x}_i\in D^*, \mathbf{x}_I\neq\mathbf{x}_i}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)c}{d(\mathbf{x}_0,\mathbf{x}_i)}}{\frac{1}{d(\mathbf{x}_*,\mathbf{x}_I)} + \sum_{i=1:\mathbf{x}_i\in D^*, \mathbf{x}_I\neq\mathbf{x}_i}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)}{d(\mathbf{x}_*,\mathbf{x}_i)} + \sum_{i=1:\mathbf{x}_i\in \overline{D^*}}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x}_*)}{d(\mathbf{x}_*,\mathbf{x}_i)}} > c$$

Then we know that when $n > N_0$, $\hat{f}_n(\mathbf{x}_*) > c$. By Lemma 1, $\mathbf{x}_*$ is a limit point of $\{\mathbf{x}_i\}_{i=1}^{\infty}$. □

Lemma 3. *If $\mathbf{x}_0$ is a limit point of $\{\mathbf{x}_i\}_{i=1}^{\infty}$, then there exists $\delta_{\mathbf{x}_0} > 0$ such that for all $\mathbf{x} \in O_{\delta_{\mathbf{x}_0}}(\mathbf{x}_0)$, $\mathbf{x}$ is a limit point.*

Proof. Consider $\delta_0 > 0$ such that $O_{4\delta_0}(\mathbf{x}_0) \subset D$. Since $\mathbf{x}_0$ is a limit point, we can find $\{\mathbf{x}_{n_l}\}_{l=1}^{\infty} \to \mathbf{x}_0$. Then let $N_{\mathbf{x}_0} = \min\{n_l : \sum_{i=1}^{n_l} I_{d(\mathbf{x}_0,\mathbf{x}_i)<\delta_0}\}$, where $I_{d(\mathbf{x}_0,\mathbf{x}_i)<\delta_0}$ indicates whether $\mathbf{x}_i$ lies within open ball $O_{\delta_0}(\mathbf{x}_0)$. Then for all $\mathbf{x} \in O_{\delta_0}(\mathbf{x}_0)$, when $n > N_{\mathbf{x}_0}$, all the nearest k sampled points to $\mathbf{x}$ will be inside of $O_{4\delta_0}(\mathbf{x}_0) \subset D$. Thus we have $\hat{f}_n(\mathbf{x}) > c$, for all $n > N_{\mathbf{x}_0}$. □

The δ_0 in Lemma 3 is not necessarily small. Actually, δ_0 can be $\frac{\delta_{max}}{4}, \delta_{max} = sup\{\eta, O_\eta(\mathbf{x}_0) \subset D\}$.

With above lemmas, we can begin to prove Theorem 6. For $\mathbf{x}_0 \in D$, let $\mathbf{x}_*$ be the limit point identified in Lemma 2, since D is path connected, we can find a path $\{\phi(t) \in D, \forall t_{\mathbf{x}_*} \leq t \leq t_{\mathbf{x}_0}\}$ which connect $\mathbf{x}_*$ and $\mathbf{x}_0$, i.e., $\phi(t_{\mathbf{x}_*}) = \mathbf{x}_*, \phi(t_{\mathbf{x}_0}) = \mathbf{x}_0$. Then by Lemma 3, we can find $O_{\delta_{\mathbf{x}_*}}(\mathbf{x}_*)$ such that for all $\mathbf{x} \in O_{\delta_{\mathbf{x}_*}}(\mathbf{x}_*)$, $\mathbf{x}$ is a limit

point. Then we can find $\mathbf{x}_1 = \partial O_{\frac{\delta_{\mathbf{x}*}}{2}}(\mathbf{x}_*) \cap \phi(t)$ and construct a open ball around $\mathbf{x}_1$. We can do this again and again until there is an open ball around $\mathbf{x}_k$ including $\mathbf{x}_0$ and thus $\mathbf{x}_0$ is a limit point. □

Theorem 7. *Under the same condition of Theorem 6,* $\forall \mathbf{x} \in \partial D$, $\mathbf{x}$ *is a limit point.*

Proof. For all $\mathbf{x} \in \partial D$ and for $m \in \mathbb{N}$, the set of natural numbers, we can find a point $\mathbf{x}_{\frac{1}{m}} \in D \cap O_{\frac{1}{m}}(\mathbf{x})$. Thus $\mathbf{x}_{\frac{1}{m}} \in D$ and therefore from Theorem 6, it is a limit point. Then we can find $\mathbf{x}_{n_m} \in \{\mathbf{x}_i\}_{i=1}^{\infty}$ satisfying $n_m > n_{m-1}$ and $\mathbf{x}_{n_m} \in O_{\frac{1}{m^2}}(\mathbf{x}_{n_m}) \cap D$. Then $\{x_{n_m}\}_{m=1}^{\infty}$ is a subsequence converging to $\mathbf{x}$. □

Remark 5. *Theorem 6 and Theorem 7 can be interpreted as follows: every point inside or on the boundary of the feasible region is a limit point of the design sequence* $\{\mathbf{x}_i\}_{i=1}^{\infty}$. *This means that the design sequence is dense within the feasible region. We also find that the sequential strategy is robust to the shape of the feasible region. Path connection of the feasible region is the only shape restriction necessary for the results to hold.*

We also find that the initial design is important for the above theorem to hold. The theorems require that the ***initial design should have at least one point in the feasible region****. If this requirement is not satisfied, the sequential design can possibly miss the entire feasible region. This motivates us to use a initial design with good space-filling properties.*

In case the feasible region is not single connected, as long as (i) the feasible region is union of different path connected sets and (ii) the initial design generates points in every set, it is easy to extend the above theorems to the following corollary.

Corollary 1. *If the feasible region D can be expressed as the union of path connected sets, i.e. $D = \bigcup_{j=1}^{m} D_j$, where D_j is path connected set and there exists initial design point $\mathbf{x}_{I,j} \in \{\mathbf{x}_i\}_{i=1}^{\infty}$ for $j = 1, \cdots, m$ such that $\mathbf{x}_{I,j} \in D_j$, then $\forall \mathbf{x} \in D^*$, $\mathbf{x}$ is a limit point.*

Theorem 8. *For all $\epsilon > 0, \forall \mathbf{x} \in D \cap B(\partial D, \epsilon)^C$, then $\exists N, \forall n > N$, $\hat{f}_n(\mathbf{x}) > c$.*

Similarly, for all $\epsilon > 0, \forall \mathbf{x} \in B(D, \epsilon)^C$, then $\exists N, \forall n > N$, $\hat{f}_n(\mathbf{x}) < c$.

Proof. For all $\mathbf{x} \in D \cap B(\partial D, \epsilon)^C$, since $\mathbf{x}$ is a limit point, we can find $\{\mathbf{x}_{n_l}\}_{l=1}^{\infty}$ converges to $\mathbf{x}$. Since $\mathbf{x} \in D$, we can find $O_{4\delta_{\mathbf{x}}}(\mathbf{x}) \subset D$. Let $N_{\mathbf{x}} = \min\{n_l : \sum_{\mathbf{x}_i \in \{\mathbf{x}_{n_l}\}_{l=1}^{\infty}, i \leq n_l} I_{d(\mathbf{x}, \mathbf{x}_i) < \delta_{\mathbf{x}}}\}$. Then when $n > N_{\mathbf{x}}$, for all $\mathbf{x}_0 \in O_{\delta_{\mathbf{x}}}(\mathbf{x})$, the k-NN neighborhood of $\mathbf{x}_0$ only includes points in the feasible region, and thus $\hat{f}_n(\mathbf{x}_0) > c$.

Since $D \cap B(\partial D, \epsilon)^C$ is a closed bounded set, there will be finite number of balls, $O_{\delta_{\mathbf{x}_1}}(\mathbf{x}_1), ..., O_{\delta_{\mathbf{x}_m}}(\mathbf{x}_m)$ which cover $D \cap B(\partial D, \epsilon)^C$. Let $N = max\{N_{\mathbf{x}_1}, ..., N_{\mathbf{x}_m}\}$, then for all $\mathbf{x} \in D \cap B(\partial D, \epsilon)^C$, $\hat{f}_n(\mathbf{x}) > c$ when $n > N$.

The second part of the theorem can come from similar argument. □

Remark 6. *This theorem can be interpreted as: the k-NN estimation method will provide accurate classification between feasible and infeasible region except in small area near the boundary of feasible region when the design sequence are generated.*

This theorem also introduces two different types of error in estimation as below:

Type I error $f(\mathbf{x}) > c$ and $\hat{f}_n(\mathbf{x}) < c$.

Type II error $f(\mathbf{x}) < c$ and $\hat{f}_n(\mathbf{x}) > c$.

The first type of error means we have mistakenly estimate a point in the feasible region to be in the infeasible region. The second one means we estimate a point in

the infeasible region to be in the feasible region. Since our goal is to carve out the infeasible region, the Type II error is usually more important in reality.

Another property related to the accuracy of k-NN estimation in the feasible region comes from the following theorem:

Theorem 9. *If $f(\mathbf{x})$ is a continuous function in Ω, then*

$$\lim_{n\to\infty} \max_{\mathbf{x}\in D^*} |f(\mathbf{x}) - \hat{f}_n(\mathbf{x})| = 0 \tag{4.10}$$

Proof. We will prove for all $\epsilon > 0$, there exists N such that $\forall n > N$, $\forall \mathbf{x} \in D^*$,

$$|f(\mathbf{x}) - \hat{f}_n(\mathbf{x})| < \epsilon$$

For $\mathbf{x} \in D^*$, since $f(\mathbf{x})$ is a continuous function, we can find $O_{4\eta_{\mathbf{x}}}(\mathbf{x}) \subset D$ such that $\forall \mathbf{x}_1, \mathbf{x}_2 \in O_{4\eta_{\mathbf{x}}}(\mathbf{x}), |f(\mathbf{x}_1) - f(\mathbf{x}_2)| < \frac{\epsilon}{2}$.

Moreover, since $\mathbf{x}$ is a limit point, we can find a subsequence $\{\mathbf{x}_{n_l}\}_{l=1}^{\infty}$ converge to $\mathbf{x}$. Let $N_{\mathbf{x}} = \min\{n_l : \sum_{\mathbf{x}_i \in \{\mathbf{x}_{n_l}\}_{l=1}^{\infty}, i \le n_l} I_{d(\mathbf{x},\mathbf{x}_i)<\eta_{\mathbf{x}}}\}$ be the index while at least k points from $\{\mathbf{x}_{n_l}\}_{l=1}^{\infty}$ are inside of $O_{\eta_{\mathbf{x}}}$. Then when $n > N_{\mathbf{x}}$, we have for all $\tilde{\mathbf{x}} \in O_{\eta_{\mathbf{x}}}$, the k-NN neighborhood of $\tilde{\mathbf{x}}$ is contained in $O_{4\eta_{\mathbf{x}}}$. Then we have:

$$\begin{aligned}
|\hat{f}_n(\tilde{\mathbf{x}}) - f(\tilde{\mathbf{x}})| &\le |\hat{f}_n(\tilde{\mathbf{x}}) - f(\mathbf{x})| + |f(\mathbf{x}) - f(\tilde{\mathbf{x}})| \\
&= \Big|\frac{\sum_{i=1}^{n} \frac{I_k(\mathbf{x}_i,\tilde{\mathbf{x}}) f(\mathbf{x}_i)}{d\left(\tilde{\mathbf{x}},\mathbf{x}^{(i)}\right)}}{\sum_{i=1}^{n} \frac{I_k(\mathbf{x}_i,\tilde{\mathbf{x}})}{d(\tilde{\mathbf{x}},\mathbf{x}_i)}} - f(\mathbf{x})\Big| + |f(\mathbf{x}) - f(\tilde{\mathbf{x}})| \\
&\le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon
\end{aligned}$$

Then we know for each points $\mathbf{x}$ in D^*, there exists $\eta_{\mathbf{x}}$ and $N_{\mathbf{x}}$, such that

$$max_{\tilde{\mathbf{x}} \in O_{\eta_{\mathbf{x}}} \cap D} |\hat{f}_n(\tilde{\mathbf{x}}) - f(\tilde{\mathbf{x}})| \leq \epsilon$$

Then by finite selection theorem, we can select finite number of open balls which can cover D^*, which lead to the theorem. □

Remark 7. *This theorem can be interpreted as: when the points are sequentially selected, the estimation of the response surface within feasible region is consistent with the true value. This is important since our ultimate goal is to obtain a precise estimation of the response surface in feasible region. The result does not hold for the infeasible region.*

However, the estimation provided by k-NN IDW may not be an optimal estimator even though its consistency. In the next Chapter, another algorithm will be proposed to generate precise estimation and optimal design simultaneously.

4.6 Adjusting the algorithm to reduce prediction bias, selecting optimal k and choosing initial design

4.6.1 Adjustment of k-NN estimator

Even though k-NN IDW is more robust than original IDW, it is still sensitive to distribution of sampled points. As the points are sequentially selected, the distribution of points between feasible region and infeasible region becomes more different as

the infeasible region is carved out. Consequently, the predicted response at points near the boundary of the feasible region is likely to be biased. Clearly, the k-NN neighborhood of such points is likely to contain more points in the feasible region compared to the infeasible region, resulting in an overestimation of the unsampled points near boundary.

However, in reality, it is difficult to find exact distribution of $\{\mathbf{x}_i\}_{i=1}^{\infty}$. So it will be impossible to directly adjust the weights. However, a simple weight adjusted k-NN estimator (k-NN IDW-adj) can be proposed to balance the weights if there are points within the k-NN neighborhood lying in both the feasible and infeasible region.

If k-NN neighborhood of $\mathbf{x}$ consists of m points in D^* and $k-m$ points in $\overline{D^*}$ and $m > \frac{k}{2}$. Then we can estimate $\hat{f}_n(\mathbf{x})$ as:

$$\hat{f}_n(\mathbf{x}) = \frac{\frac{1}{m}\sum_{i=1,\mathbf{x}_i\in D^*}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x})f(\mathbf{x}_i)}{d(\mathbf{x},\mathbf{x}_i)} + \frac{1}{k-m}\sum_{i=1,\mathbf{x}_i\in \overline{D^*}}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x})f(\mathbf{x}_i)}{d(\mathbf{x},\mathbf{x}_i)}}{\frac{1}{m}\sum_{i=1,\mathbf{x}_i\in D^*}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x})}{d(\mathbf{x},\mathbf{x}_i)} + \frac{1}{k-m}\sum_{i=1,\mathbf{x}_i\in \overline{D^*}}^{n} \frac{I_k(\mathbf{x}_i,\mathbf{x})}{d(\mathbf{x},\mathbf{x}_i)}}. \tag{4.11}$$

The idea of adjustment is to replicate points in the infeasible region to generate a simple random sample in k-NN neighborhood of $\mathbf{x}$. Moreover, since k is a pre-defined integer, this adjustment will not affect the theoretical results in previous session. All the theorems in Section 4.5 will still hold. The space-filling strategy based on adjustment of IDW is referred as IDW-MM-adj henceforth.

4.6.2 Selection of k

In k-NN estimation, the selection of k is very crucial. A small k leads to inaccurate estimation whereas a large k makes the interpolation more volatile to the distribution

of sampled points.

Here we propose an adaptive version of the IDW-MM algorithm based on an m-fold stratified cross validation method (Picard and Cook, 1984), which can be used to select the optimal k in the process of sampling. In the process of sampling, different checkpoints are set, for example, every 10 points. At each check point, the stratified cross validation is used to select the optimal k based on the points already sampled. Then this value of k is used to select new points until the next checkpoint.

At each checkpoint, the m-fold cross validation algorithm can be described as follows: suppose we have already sampled n points, among which n_1 points are in the feasible region. The n_1 points in the feasible region and $n-n_1$ points in the infeasible region are randomly assigned to m groups. Then for the i^{th} group, $i = 1, \cdots m$, we use points from all groups except the i^{th} group to obtain k-NN estimates for the response at design points in the i^{th} group.

Type I and Type II errors for the i^{th} group can be easily obtained. Let $er_{i,1}(k), er_{i,2}(k)$ denote the Type I and Type II errors in m-fold cross validation corresponding to the points in feasible region in $i^{th}, i = 1, \cdots, m$ group with the value k. Then we can select the optimum k based on the following criterion function:

$$p(k) = \lambda \sum_{i=1}^{m} er_{i,1}(k) + (1-\lambda) \sum_{i=1}^{m} er_{i,2}(k). \tag{4.12}$$

The weights λ can be chosen between 0 and 1 in the above criterion function. Since Type II error is generally more important than Type I error, here we select $\lambda = 0.2$.

4.6.3 Choice of initial design

The choice of initial design is important for the proposed algorithm, as discussed in Remark 5. The initial design serves the task of exploration of the whole design space Ω. We use Latin hypercube design (LHD) (McKay et al., 1979) of a fixed run-size n_0 as initial design in the proposed algorithm. Latin hypercube designs generate distributions of plausible collections of parameter values from a multidimensional distribution. It is a generalization of Latin square to an arbitrary number of dimensions. Different extensions to Latin hypercube design have been proposed. For example, Tang (1993) proposed orthogonal array-based Latin hypercube design. Joseph and Hung (2008) considered and discussed orthogonal maximin Latin hypercube design.

Computer codes which can generate Latin hypercube design are well established in different computer languages. The 'lhs' package in R has been used to generate initial design. Two examples of LHD design with 16 points on $[0, 1]^2$ can be found in Figure 4.4.

4.7 Performance of algorithms for deterministic function

In this section, the performance of IDW-MM is evaluated for a certain deterministic response function. The following measures are proposed to evaluate the performance of the algorithm:

1. Expected percentage of points in the feasible region, defined as $P = \mathbb{E}\left(\frac{\sum_{i=1}^{n} I_{\mathbf{x}\in D^*}}{n}\right)$,

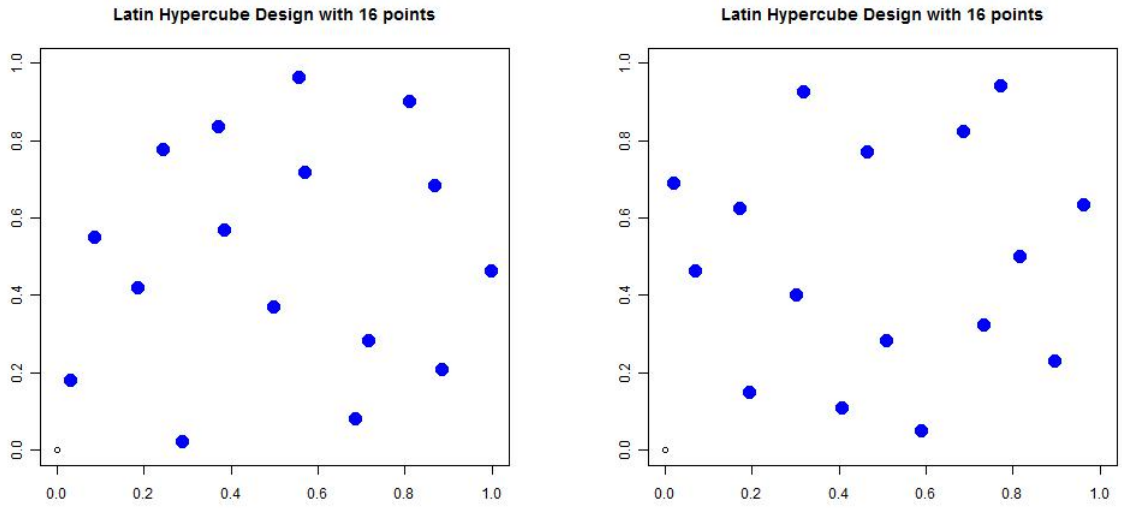

Figure 4.4: Two Latin Hypercube Designs with 16 points on $[0,1] \times [0,1]$.

where the expectation is taken with respect to the random choice of the initial design.

2. Ratio of minimum distance between sample in feasible region to minimum distance between sample obtained by sequential maximin design only in feasible region with same number of points, defined by $RM = \mathbb{E}\left(\frac{min_{\mathbf{x}_i,\mathbf{x}_j \in \mathscr{D}_{SEQ}} d(\mathbf{x}_i,\mathbf{x}_j)}{min_{\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j \in \mathscr{D}_{GREEDY}} d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)}\right)$. Design points set $\mathscr{D}_{SEQ} = \{\mathbf{x}_i\}_{i=1}^m$ represent points generated by IDW-MM/IDW-MM-adj in the feasible region D^*. Design points set $\mathscr{D}_{GREEDY} = \{\tilde{\mathbf{x}}_i\}_{i=1}^m$ are sequential space-filling design on D^* generated by maximin greedy algorithm (4.6) with $m = \sum_{i=1}^n I_{\mathbf{x}_i \in D^*}$.

3. Ratio of average of inverse distance between sample in the feasible region to average of inverse distance between sample obtained by sequential maximin design only in feasible region with the same number of points, defined by $RA =$

$$\mathbb{E}\left(\left(\frac{\sum_{\mathbf{x}_i,\mathbf{x}_j\in\mathscr{D}_{SEQ}}\frac{1}{d(\mathbf{x}_i,\mathbf{x}_j)^2}}{\sum_{\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j\in\mathscr{D}_{GREEDY}}\frac{1}{d(\tilde{\mathbf{x}}_i,\tilde{\mathbf{x}}_j)^2}}\right)^{\frac{1}{2}}\right).$$

We want P to be large since it reflects the percentage of points in the feasible region. It measures whether the algorithm can successfully carve out the infeasible region. We also want RM to be large and RA to small. Both RM and RA reflect whether space-filling properties of the algorithms are properly maintained in comparison to space-filling design generated by (4.6) within the feasible region.

We evaluate the performance of proposed space-filling algorithm with a standard test function widely used in global optimization literature. Branin function (Branin, 1972), which is a popular test function used to evaluate the performance of optimization algorithm will be utilized. The function has three global minima. Because our objective is to generate a space-filling design within a feasible region rather than global optimization, we use the following transformations of Branin function as in Dasgupta (2007):

Let $B(x, y)$ denote the negative of the Branin function such that

$$B(x,y) = -\left((y - 5x^2/4\pi^2 + 5x/\pi - 6)^2 + 10(1 - 1/8\pi)cos(x) + 10\right), -5 \le x \le 10, 0 \le y \le 15.$$

According to Dasgupta (2007), the above function achieves maximum (-0.3979) at $(-\pi, 12.25), (\pi, 2.25)$ and (9.4248,2.25). The minimum is around 305. We transform B to a function g with a domain $[0,1]\times[0,1]$ and range $[0,1]$ by applying the following transformation:

$$g(u,v) = \frac{B\left(15u - 5, 15v\right) + 305}{-.3979 + 305}, 0 \le u \le 1, 0 \le v \le 1.$$

Finally, another transformation yields a large non-convex region with zero yields:

$$p(u,v) = max\{\frac{g(u,v) - .8}{.2}, 0\}, 0 \leq u \leq 1, 0 \leq v \leq 1. \tag{4.13}$$

The contour plot of the adjusted Branin function $p(u,v)$ is shown in Figure 4.5.

Two different versions of space-filling algorithms are used in simulation: (i) IDW-MM and (ii) IDW-MM-adj. We use two different thresholds to test out algorithm: (i) $c = 0.2$ and (ii) $c = 0.7$. Note that for $c = 0.2$, the feasible region is single connected; however, for $c = 0.7$, it is union of three single connected sets. The sequence of design points generated by IDW-MM and IDW-MM-adj with $k = 5$ and the two thresholds are shown in Figure 4.6. We find IDW-MM and IDW-MM-adj effectively carves out the infeasible region and possesses space-filling properties. Moreover, although for $c = 0.7$, the feasible region consists of three isolated single connected parts, the proposed algorithm still works.

Intensive simulation has been done to evaluate the performance of IDW-MM and IDW-MM-adj with respect to different choice of k. For each algorithm, repeated simulations (500 times) with different initial Latin hypercube Design and k have been implemented. In each simulation, IDW-MM and IDW-MM-adj sequentially generated 100 design points after 16 initial design points so that $n^* = 116$. The results reported represent the median of each performance measure over repeated simulations and are shown in Figure 4.7.

From Figure 4.7, we observe that for IDW-MM, $k = 3$ for $c = 0.2$ and $k = 3$ for $c = 0.7$ can lead to highest percentage of points in the feasible region. For IDW-MM-

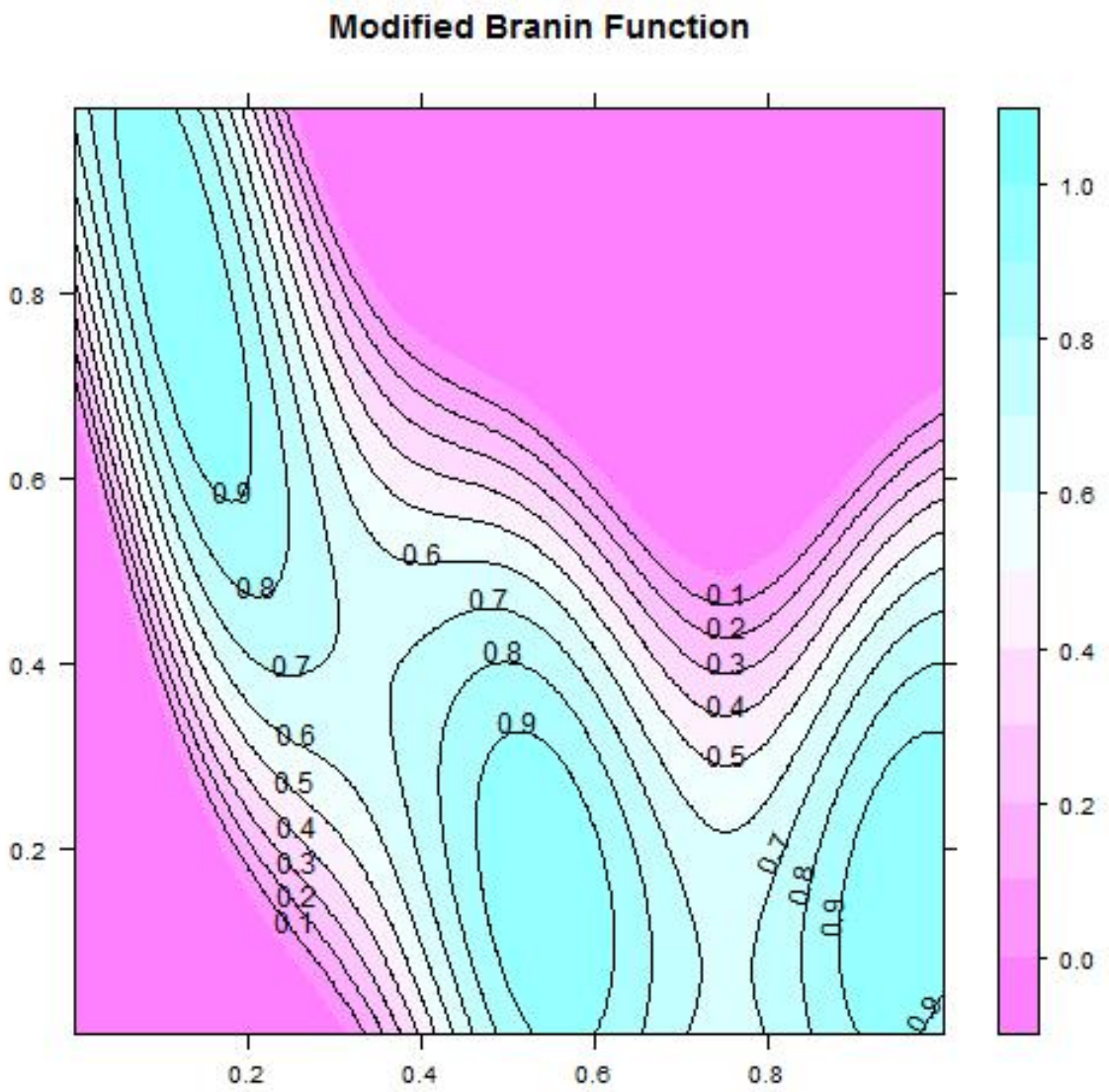

Figure 4.5: Contour Plot of Modified Branin Function

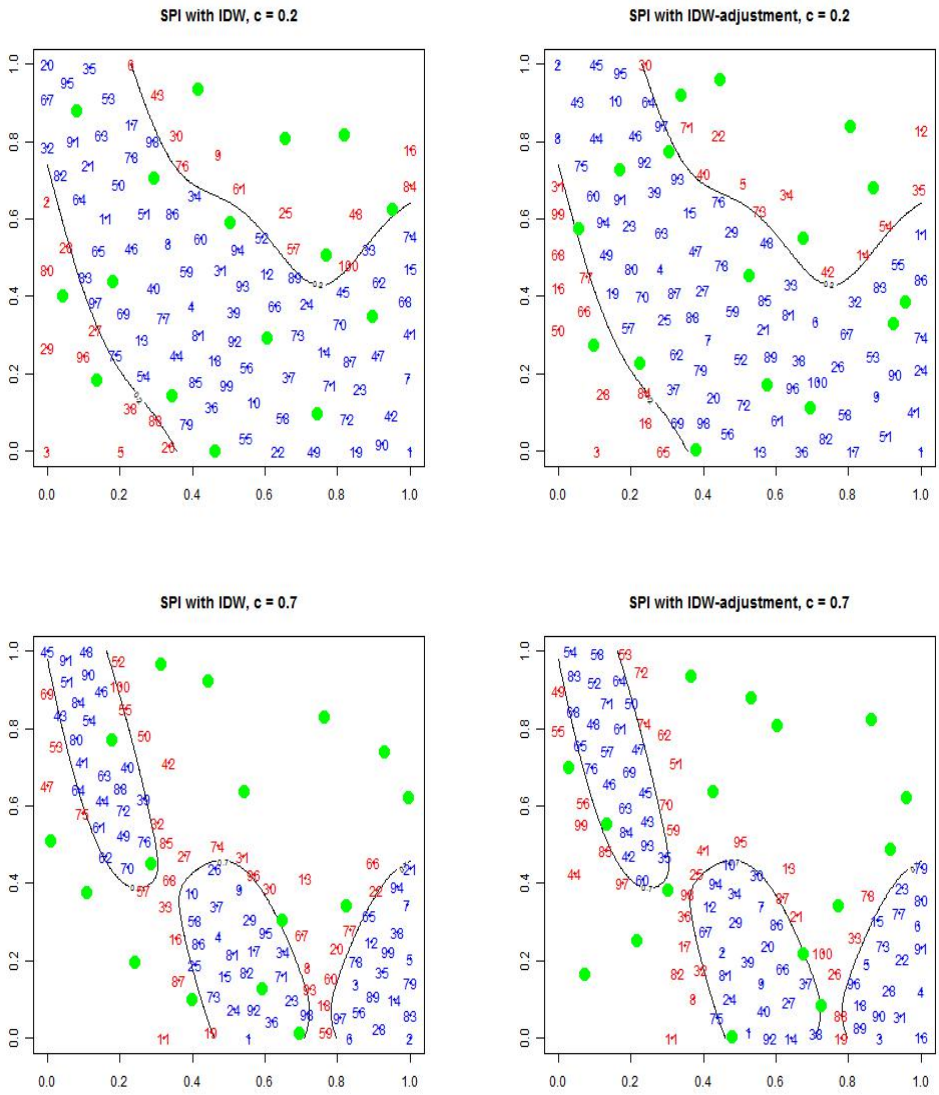

Figure 4.6: Example of proposed algorithm for two thresholds. The left one is with threshold 0.2 and right one is with 0.7. The green points are initial design points. The blue and red points are sequential design points in feasible and infeasible region.

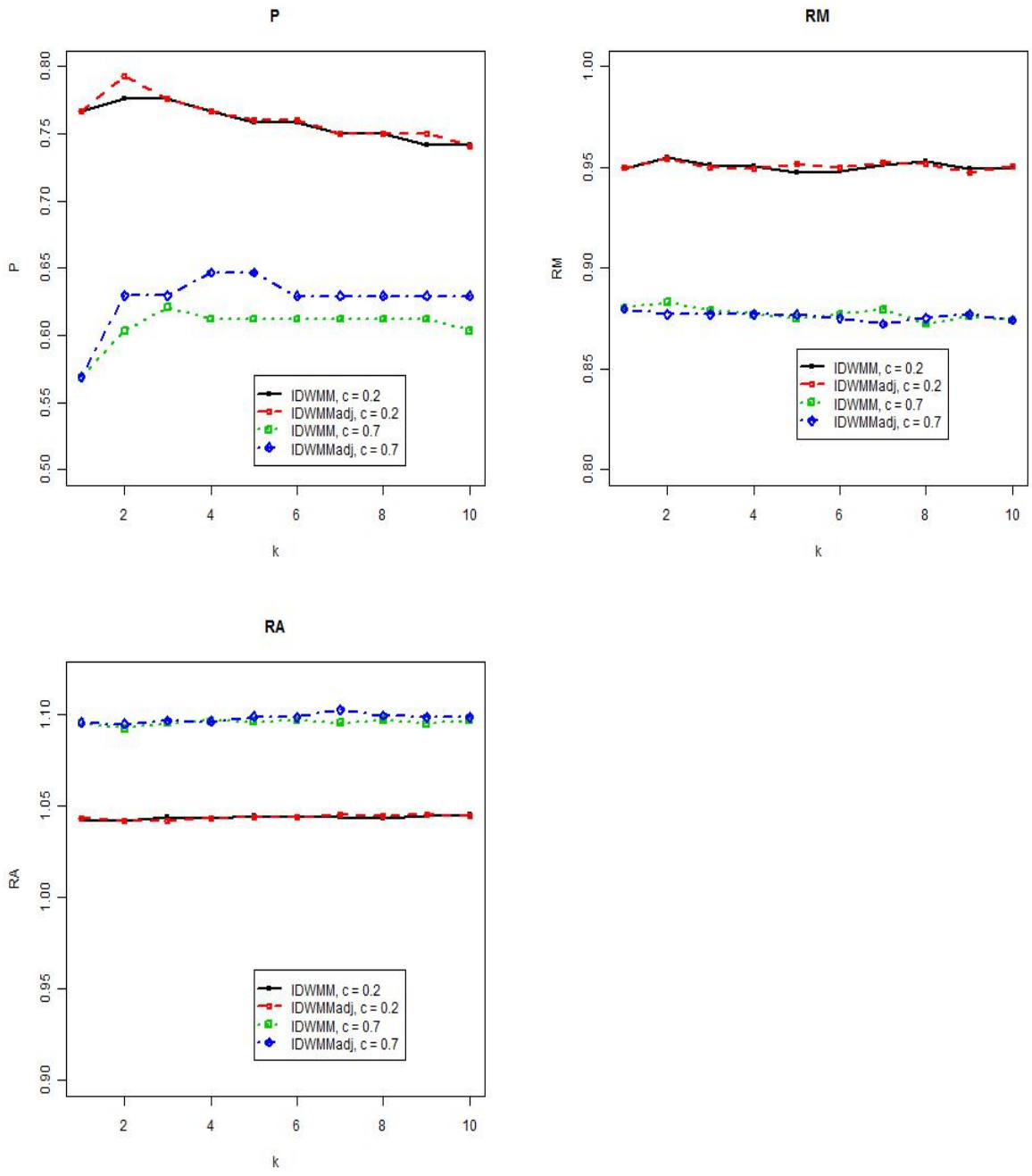

Figure 4.7: Performance of P, RM and RA for IDW-MM and IDW-MM-adj with respect to different k.

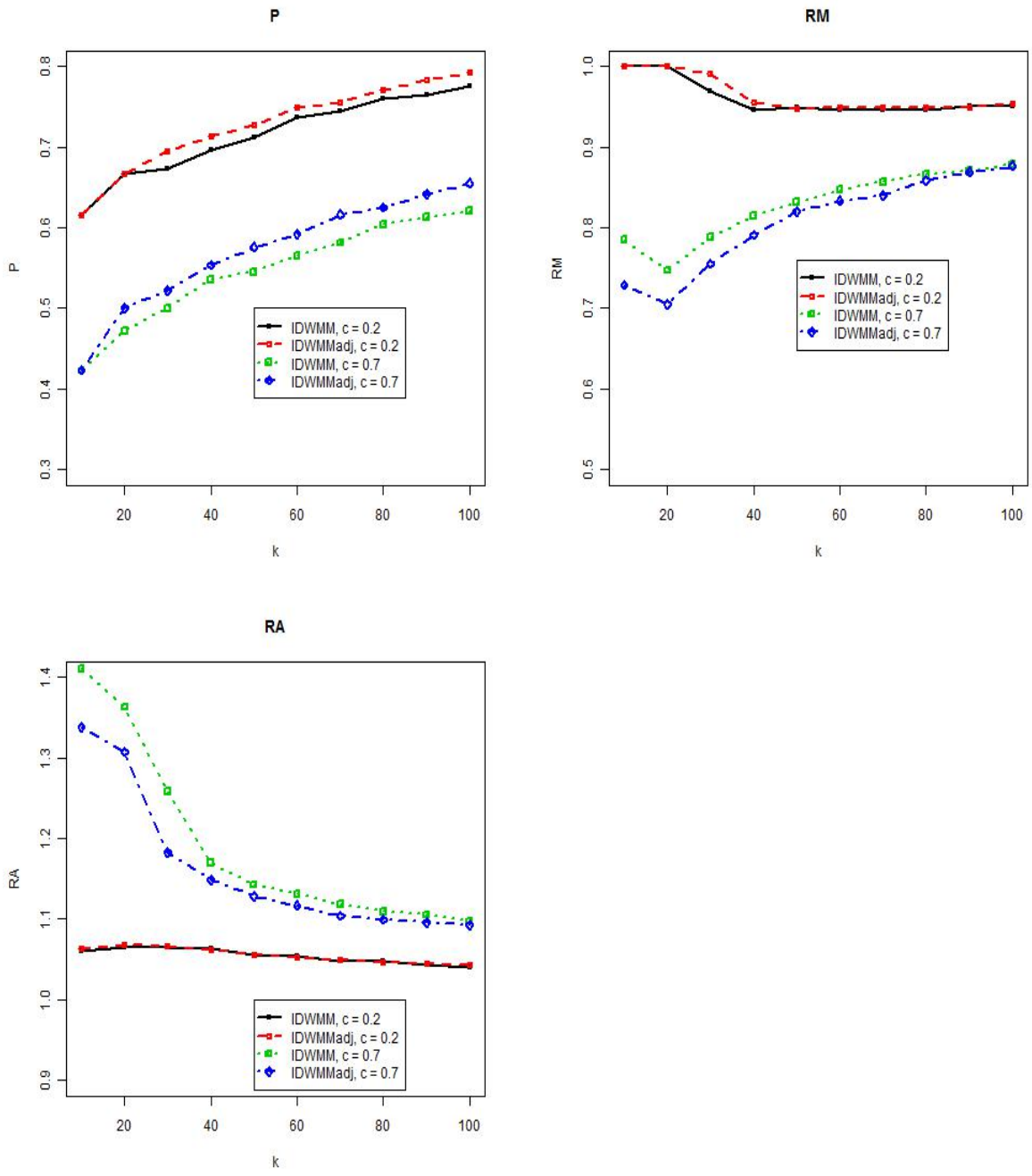

Figure 4.8: Performance of P, RM and RA for IDW-MM and IDW-MM-adj with optimal k and different sequential design size.

adj, $k = 2$ and $k = 4$ can lead to highest percentage of points in the feasible region for the thresholds $c = 0.2$ and $c = 0.7$ respectively. We can find the percentage of points achieve maximum for some k between 1 and 10. When k is small, the prediction is not accurate due to the small number of data in the k-NN neighborhood. When k is large, the distribution of the sampled points lead to bias in prediction. Given the optimal k for two thresholds, performance of IDW-MM and IDW-MM-adj with respect to different size of design are shown in Figure 4.8, from which the following conclusions can be obtained:

1. IDW-MM-adj leads to higher percentage points in the feasible region than IDW-MM. We observe that the adjusted version of the k-NN IDW method provides more accurate estimation.

2. When $c = 0.2$, RM and RA are very close to 1 for both IDW-MM and IDW-MM-adj. This means space-filling property within the feasible region is properly maintained. When $c = 0.7$, the space-filling property is worse than the case of $c = 0.2$.

3. From Figure 4.8, when $c = 0.2$, the performance measures of IDW-MM and IDW-MM-adj stabilize very quickly with respect to the size of design. IDW-MM and IDW-MM-adj can achieve satisfactory space-filling property as well as percentage of points in the feasible region very quickly. When $c = 0.7$, the number of runs to achieve stability is much larger.

4. No significant difference in RM and RA is observed between IDW-MM and IDW-MM-adj, for both thresholds. This means that while IDW-MM results in

possible bias in estimation, it does not influence the space-filling properties.

Finally we have implemented the algorithm with cross validation. In cross validation, the value k is not required to be specified. The sequential design still consists of 100 points after 16 initial design points. The cross validation is conducted every 10 points. Repeated simulations are used to evaluate the performance of cross validation. Performance measures P, RM, RA and the optimal choice of k at each check points are shown in Figure 4.9 and Figure 4.10.

We find that the cross validation can successfully identify the optimal k as well as select the design points simultaneously. The choice of k is more or less in agreement with the optimal k chosen earlier from Figure 4.7. Moreover, the performance of the algorithm with cross validation is very similar to those generated by IDW-MM and IDW-MM-adj with optimal k in Figure 4.8.

4.8 Summaries and Conclusion

Two space-filling strategies, IDW-MM and IDW-MM-adj are proposed in this Chapter. The algorithms are seen to successfully carve out the infeasible region and sequentially generate new points with space-filling property in the feasible region. Convergence properties of the algorithm has been established and simulations has been conducted to evaluate the algorithm. However, in spite of the above mentioned advantages, the proposed space-filling design has the following weaknesses:

1. The function evaluation step of the algorithm suffers from the inherited weakness of the IDW predictor as discussed earlier. Use of a Gaussian process model

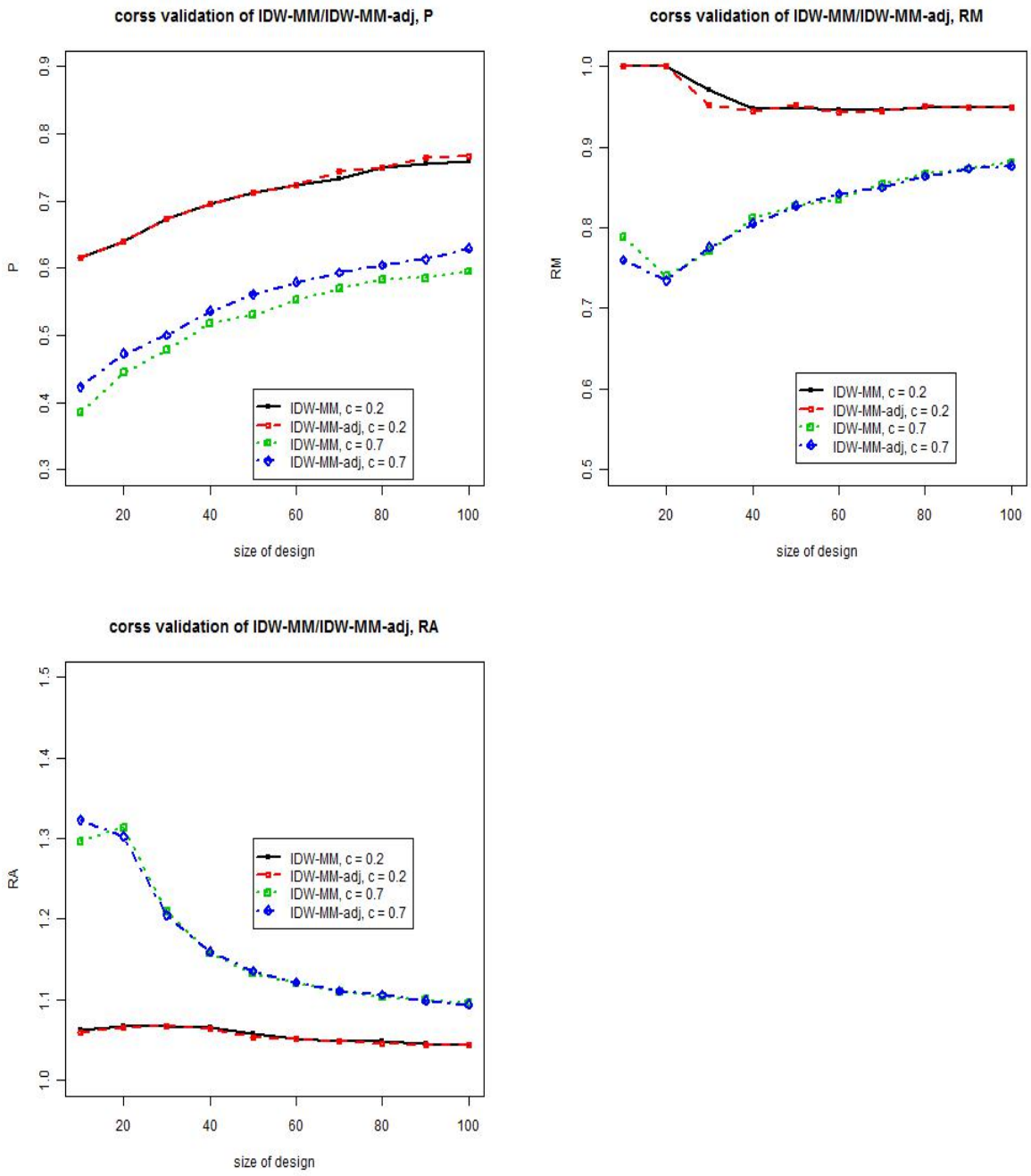

Figure 4.9: Performance of P, RM and RA for IDW-MM and IDW-MM-adj with cross validation.

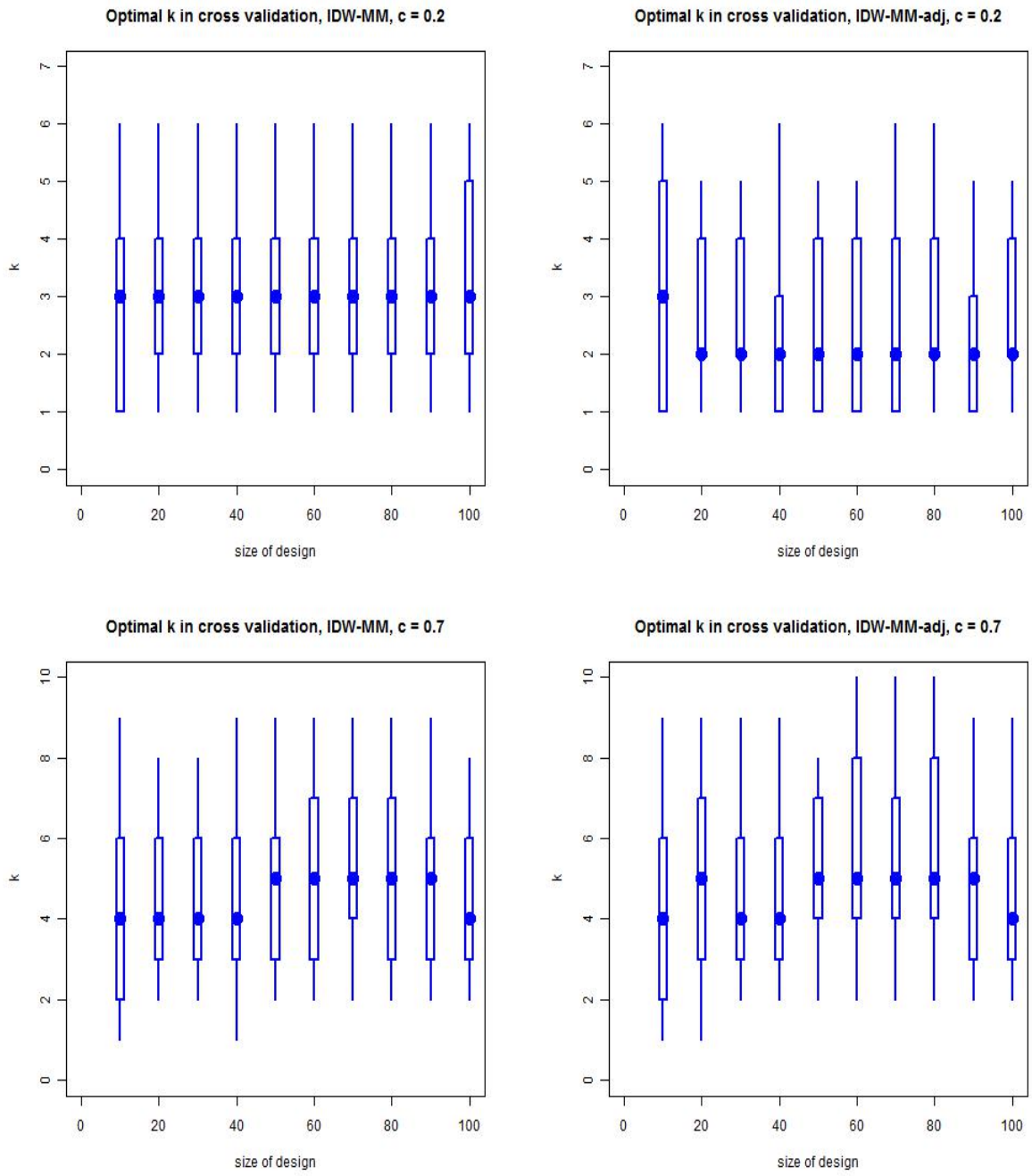

Figure 4.10: Optimal k at each check points in cross validation. The blue points represent median of chosen k. The rectangle box represent 50% confidence intervals and lines represent 95% confidence interval.

based prediction approach may help the experimenter achieve improved prediction performance.

2. In the second stage of the algorithm, we have used maximin criteria to select the subsequent points. However, the proposed algorithm is heuristic in nature and may not necessarily guarantee a true maximin design.

3. The proposed version of the algorithm is grid-based. The computational cost required would be extremely high when the algorithm is applied to high dimensional response functions.

In the next chapter, a sequential strategy based on Gaussian process model will be proposed and discussed.

Chapter 5

The GP-IMSE and GP-MMSE algorithm for deterministic functions

5.1 Introduction

In Chapter 4, IDW-MM has been proposed to obtain space filling design within a 'pre-defined' feasible region within a design space with a complex response surface. The sequential design includes two part (i) function evaluation of the response surface and (ii) selection of the next design points. The k-NN inverse distance interpolator has been used for function evaluation and maximin criterion has been used to select the next design point. The sequential strategy shows nice convergence properties and performs well in simulation study.

Two advantages of IDW are its simplicity, computational advantage and robust-

ness. However, its principle drawback is its poor prediction power (Joseph and Kang, 2011). Although some of the problems can be circumvented by using k-NN IDW, the need to choose optimal k increases the computational load.

Selection criterion for the next design point should be specified after the response surface has been estimated. In Chapter 4, maximin criterion has been used. The maximin criterion is model independent. However, when a model is used to fit the response surface, it is reasonable to use a model-based criterion for the selection of subsequent points.

The overall approach proposed in this Chapter is motivated by efficient global optimization algorithm based on Gaussian process model. Two criteria that are based on minimizing the prediction errors are proposed for iteratively selecting new design points.

This Chapter is organized as follows: in Section 5.2 an introduction about Gaussian process model is given. Then we briefly introduce the efficient global optimization (EGO) algorithm proposed by Jones et al. (1998) in Section 5.3. After that, two sequential strategies inspired by EGO are proposed and discussed. Simulations are used to evaluate the performance of proposed strategies, and to compare them to the k-NN IDW and other competing algorithms.

5.2 Gaussian process model for deterministic response function and Kriging

In Gaussian process modeling, the deterministic response function is assumed to be a realization from a Gaussian random function. A function $y(\mathbf{x}), \mathbf{x} \in \Omega$ is a Gaussian random function if for any choice of $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$, the vector $(y(\mathbf{x}_1), y(\mathbf{x}_2), \cdots, y(\mathbf{x}_n))$ has a multivariate normal distribution (Santner et al., 2003). Then the function is determined by their mean function $\mathbb{E}(y(\mathbf{x}))$ and covariance function:

$$C(\mathbf{x}_1, \mathbf{x}_2) = Cov(\mathbf{x}_1, \mathbf{x}_2) \tag{5.1}$$

Given the responses obtained from sampled points, prediction of response from unsampled points can be obtained using *Kriging method* (Santner et al., 2003). The intuition of Kriging method can be explained as follows: since the response follows a multivariate normal distribution, the mean and the covariance matrix are the only two things required to completely specify the model. Once the mean and covariance structure are estimated, the best linear predictor of the response function at any arbitrary unsampled point can be obtained. From a Bayesian prospective, the Kriging method identifies the posterior distribution of the response surface given the data, if a Gaussian process prior is given.

The Kriging method was first developed by Georges Matheron, who used the method to identify the distance-weighted average gold grades at the Witwatersrand reef complex in South Africa. It has been widely used in geostatistics, for example, to estimate the rainfall in certain area (Barancourt et al., 1992). Despite its geostatistical

origin, it is also proved to be a powerful interpolation method in many field, including computer experiment.

There are many different versions of Kriging methods. The most commonly used Kriging methods are simple and universal Kriging. In simple Kriging, the mean of the response surface, $\mathbb{E}(y(\mathbf{x}))$, is assumed to be a constant. In universal Kriging, the trend of the response surface follows regression based model, which can provide better prediction but involves more computational power. Examples of more advanced Kriging methods include Kriging with qualitative and quantitative factors (Qian et al., 2008) and Limit Kriging (Joseph, 2006).

Even though different versions of Kriging methods have been proposed, simple Kriging can provide good estimation of complex response function (Jones et al., 1998). In this Chapter, simple Kriging is used as function evaluation method. A mathematical description of simple Kriging is given in the following subsection.

5.2.1 Gaussian process model for deterministic function

Let $f(\mathbf{x})$ be a deterministic response function and $\mathbf{x} = (x_1, x_2, \cdots, x_p), x_i \in [0, 1]$. Assume that n points $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ have been sampled and let the $n \times p$ matrix $\mathbf{X}_n = (\mathbf{x}_1^T, \cdots, \mathbf{x}_n^T)^T$ denote all the sampled points. We also let $\mathbf{y} = (y_1, y_2, \cdots, y_n)^T = (y(\mathbf{x}_1), y(\mathbf{x}_2), \cdots, y(\mathbf{x}_n))^T$ denote the response vector. Then the response associated with $\mathbf{x}$ is assumed to follow the following model (Santner et al., 2003):

$$y(\mathbf{x}) = \mu + \epsilon(\mathbf{x}). \tag{5.2}$$

where μ is the mean of the response surface and $\epsilon(\mathbf{x})$ is a Gaussian process with mean $\mathbb{E}(\epsilon(\mathbf{x})) = 0, Var(\epsilon(\mathbf{x})) = \sigma^2$.

The error terms in Model (5.2) for different points are assumed to be correlated. The correlation function is given by:

$$\begin{aligned} Cor\left(\epsilon(\mathbf{x}_1), \epsilon(\mathbf{x}_2)\right) &= \frac{Cov\left(\epsilon(\mathbf{x}_1), \epsilon(\mathbf{x}_2)\right)}{\sqrt{Var(\epsilon(\mathbf{x}_1)) \times Var(\epsilon(\mathbf{x}_2))}} \\ &= \frac{C(\mathbf{x}_1 - \mathbf{x}_2)}{\sigma^2} = R(\mathbf{x}_1 - \mathbf{x}_2), \end{aligned} \tag{5.3}$$

where $C(\bullet), R(\bullet)$ denote covariance and correlation functions respectively.

The covariance and correlation function should satisfy $C(\mathbf{0}) = \sigma^2, R(\mathbf{0}) = 1$. They should also be symmetric about the origin. Moreover, both $C(\bullet), R(\bullet)$ must be positive definite functions, which means for any real numbers $\omega_1, \cdots, \omega_n$ and any inputs $\mathbf{x}_1, \cdots, \mathbf{x}_n$, we must have:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \omega_i \omega_j C(\mathbf{x}_i - \mathbf{x}_j) \geq 0 \tag{5.4}$$

In Santner et al. (2003), several examples of correlation functions have been given and discussed.

One popular choice of correlation functions is the following:

$$R(\boldsymbol{h}) = exp\left(-\sum_{m=1}^{p} \theta_m |h_m|^{q_m}\right) \tag{5.5}$$

where h_m is the m^{th} element of $\boldsymbol{h}$ and $\boldsymbol{q} = (q_1, \cdots, q_p), \boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$ are pre-defined parameters.

The correlation function given by (5.5) can be interpreted as follows: if the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined as:

$$d\left(\mathbf{x}_i, \mathbf{x}_j\right) = \sum_{m=1}^{p} \theta_m |\mathbf{x}_{im} - \mathbf{x}_{jm}|^{q_m} \tag{5.6}$$

where $\mathbf{x}_{im}, \mathbf{x}_{jm}$ are m^{th} element of $\mathbf{x}_i, \mathbf{x}_j$, then the correlation between $\epsilon(\mathbf{x}_i)$ and $\epsilon(\mathbf{x}_j)$ is an exponentially decaying function of the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$.

Properties of the correlation function given by Sacks et al. (1989) have been studied by Jones et al. (1998) and Sacks et al. (1989). As noted by Sacks et al. (1989), the case $q_m = 1, m = 1, \cdots, p$ gives the product of Ornstein-Uhlenbeck process. The case $q_m = 2, m = 1, \cdots, p$ gives a process with infinitely differentiable paths. Due to its wide usage in computer experiment and intuitive interpretation, we adopt this correlation function to our problem.

5.2.2 Estimation of parameters

Model (5.2) has $2p + 2$ parameters, $\theta_1, \theta_2, \cdots, \theta_p$, $q_1, q_2, \cdots, q_p$ and μ, σ^2. The parameters can be obtained through maximum likelihood estimation (MLE). Let $\mathbf{R}(\boldsymbol{\theta}, \boldsymbol{q}, \sigma^2)$ denote the $n \times n$ matrix whose $(i, j)^{th}$ entry is $Cov\left(\epsilon(\mathbf{x}_i), \epsilon(\mathbf{x}_j)\right)$. Then the likelihood function is given by:

$$\frac{1}{(2\pi)^{\frac{n}{2}} \det(\mathbf{R}(\boldsymbol{\theta}, \boldsymbol{q}, \sigma^2))^{\frac{1}{2}}} \exp\left(-\frac{(\mathbf{y} - \mathbf{1}_n\mu)'\mathbf{R}(\boldsymbol{\theta}, \boldsymbol{q}, \sigma^2)^{-1}(\mathbf{y} - \mathbf{1}_n\mu)}{2\sigma^2}\right). \tag{5.7}$$

Given the values of $\boldsymbol{\theta}, \boldsymbol{q}$, the MLE of μ and σ^2 can be found in the following closed form:

$$\hat{\mu} = \frac{\mathbf{1}_n' \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{q}, \sigma^2)^{-1} \mathbf{y}}{\mathbf{1}_n' \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{q}, \sigma^2)^{-1} \mathbf{1}_n}. \tag{5.8}$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1}_n \hat{\mu})' \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{q}, \sigma^2)^{-1} (\mathbf{y} - \mathbf{1}_n \hat{\mu})}{n}. \tag{5.9}$$

Then by plugging (5.8) and (5.9) back into (5.7), the profile likelihood function can be obtained, maximizing which the ML estimators $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{q}}$ are obtained as follows:

$$\left(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{q}}\right) = arg \min_{(\boldsymbol{\theta}, \boldsymbol{q})} \left[-\frac{1}{2} \log \det(\mathbf{R}(\boldsymbol{\theta}, \boldsymbol{q}, \sigma^2)) \right] \tag{5.10}$$

5.2.3 The Kriging predictor

The predicted response at any unsampled point $\mathbf{x}^*$ can be computed as follows. Let $\mathbf{r}(\mathbf{x}^*)$ denote the $n \times 1$ vector of correlations between the error at $\mathbf{x}^*$ and the error terms of $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$, i.e, $\mathbf{r}_i(\mathbf{x}^*) = Cor\left(\epsilon(\mathbf{x}^*), \epsilon(\mathbf{x}_i)\right)$. Then the best linear unbiased predictor (BLUP) of $y(\mathbf{x}^*)$ (Santner et al., 2003) is given by:

$$\hat{y}(\mathbf{x}^* | \mathbf{X}_n, \mathbf{y}) = \hat{\mu} + \mathbf{r}' \hat{\mathbf{R}}^{-1} \left(\mathbf{y} - \mathbf{1}_n \hat{\mu}\right). \tag{5.11}$$

where $\hat{\mathbf{R}} = \mathbf{R}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{q}}, \hat{\sigma}^2)$ and $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{q}}, \hat{\mu}, \hat{\sigma}$ are ML estimates given by (5.8) - (5.10).

Note that the predictor given by (5.11) is an interpolator, i.e. $\hat{y}(\mathbf{x}_i | \mathbf{X}_n, \mathbf{y}) = y_i$.

The mean square error (MSE) of prediction is given by:

$$\sigma^2(\mathbf{x}^*|\mathbf{X}_n,\mathbf{y}) = \sigma^2\left[1-\mathbf{r}'\hat{\mathbf{R}}^{-1}\mathbf{r}+\frac{\left(1-\mathbf{1}'\hat{\mathbf{R}}^{-1}\mathbf{r}\right)^2}{\mathbf{1}'\hat{\mathbf{R}}^{-1}\mathbf{1}}\right]. \tag{5.12}$$

The MSE (5.12) at any of the sampled points $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n$ will be 0.

Consequently, the predicted value $\hat{y}(\mathbf{x}^*|\mathbf{X}_n,\mathbf{y})$ follows a normal distribution:

$$\hat{y}(\mathbf{x}^*|\mathbf{X}_n,\mathbf{y}) \sim N\left(y(\mathbf{x}^*|\mathbf{X}_n,\mathbf{y}),\sigma^2(\mathbf{X}_n,\mathbf{y})\right). \tag{5.13}$$

The Kriging predictor has the following advantages:

1. As mentioned earlier, Kriging method has a Bayesian interpretation. From a Bayesian prospective, the response function is assumed to be a Gaussian random function. If a prior following Gaussian process model is given to the response surface, then the Kriging method identifies the posterior distribution of response from unsampled points, given the observed data.

2. The MSE of the predictor can be easily computed. Hence confidence intervals can be conducted around the predicted values.

A disadvantage of Kriging method is that, it requires a large amount of computational power. For estimation and prediction, the inverse of correlation matrix $\hat{\mathbf{R}}$ is frequently evaluated and thus the required computational power is extremely large when the sample size is large. A detailed discussion on the comparison between the computational time required by IDW and Kriging can be found in (Joseph and Kang, 2011). The computational power cost associated with Kriging is much higher than

that cost associated with the IDW method.

5.3 Efficient Global Optimization

The efficient Global Optimization (EGO) algorithm was proposed by Jones et al. (1998). It is a sampling strategy that aims at identifying the global optimum of a deterministic function based on expected improvement criterion. EGO is an algorithm which generates a sequence of points approaching the minimum of a multi-dimensional function. Let $f(\mathbf{x})$ be the function and let $\mathbf{X}_n = \left(\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_n^T\right)^T$ denote the n sampled points, from which the response is observed. Let $f_{min} = \min\left(y_1, y_2, \cdots, y_n\right)$ be the current best function value. The Kriging method in Section 5.2.3 is used to fit the response surface and the next point $\mathbf{x}_{n+1}$ is to be selected based on this fit. Intuitively, the next point $\mathbf{x}_{n+1}$ should be selected as the one which (i) minimizes $\hat{f}(\mathbf{x})$ and (ii) maximizes the MSE $\sigma^2(\mathbf{x})$, because if the predictor at an unsampled point $\mathbf{x}$ has a larger uncertainty, it will have a larger potential to be the global minimum given the predicted value.

In Jones et al. (1998), improvement at $\mathbf{x}$ is defined as $I(\mathbf{x}) = \max(f_{min} - y(\mathbf{x}), 0)$. The improvement is a random variable because of the randomness of $y(\mathbf{x})$. Then the expected improvement at $\mathbf{x}$ is given by:

$$\mathbb{E}_y\left(I(\mathbf{x})\right) = \mathbb{E}_y\left(\max(f_{min} - y(\mathbf{x}), 0)\right). \tag{5.14}$$

Since $y(\mathbf{x}) \sim N(\hat{y}(\mathbf{x}|\mathbf{X}_n, \mathbf{y}), \sigma^2(\mathbf{x}|\mathbf{X}_n, \mathbf{y}))$, the following closed form solution for

(5.14) can be obtained, given by:

$$\mathbb{E}_y\left(I(\mathbf{x})\right) = (f_{min} - \hat{y}(\mathbf{x}))\Phi\left(\frac{f_{min} - \hat{y}(\mathbf{x})}{\sigma}\right) + \sigma\phi\left(\frac{f_{min} - \hat{y}(\mathbf{x})}{\sigma}\right). \tag{5.15}$$

where $\Phi(\bullet)$ and $\phi(\bullet)$ are the cumulative distribution function and density function of the standard normal distribution. The point with maximum expected improvement will be chosen.

EGO is a very successful global optimization algorithm. It performs very well in numerical tests conducted by Jones et al. (1998). Different versions of EGO algorithms have been discussed and compared in Jones (2001). Some follow up research has also been done on EGO algorithm. Bull (2011) has investigated the convergence rate of EGO algorithm. The idea of using expected improvement with Gaussian process model has also been widely used with different objectives. For example, Ranjan et al. (2008) used a sequential sampling strategy based on expected improvement to obtain precise estimation of contour for complicated response surface.

However, EGO is an optimization algorithm. While it can successfully identify the global optimum, it may not be able to generate space-filling design within the feasible region. If the EGO algorithm is directly used to the modified Branin function given by (4.13), the first 100 points of the sampling sequence obtained by using the EGO algorithm are shown in Figure 5.1. We observe that while the sequence of sampled points converges to optimum very quickly, the space-filling property of the sampled sequence over the feasible region is not properly maintained.

In the following section, a sequential design strategy motivated by EGO that generates a space-filling design in the feasible region will be proposed.

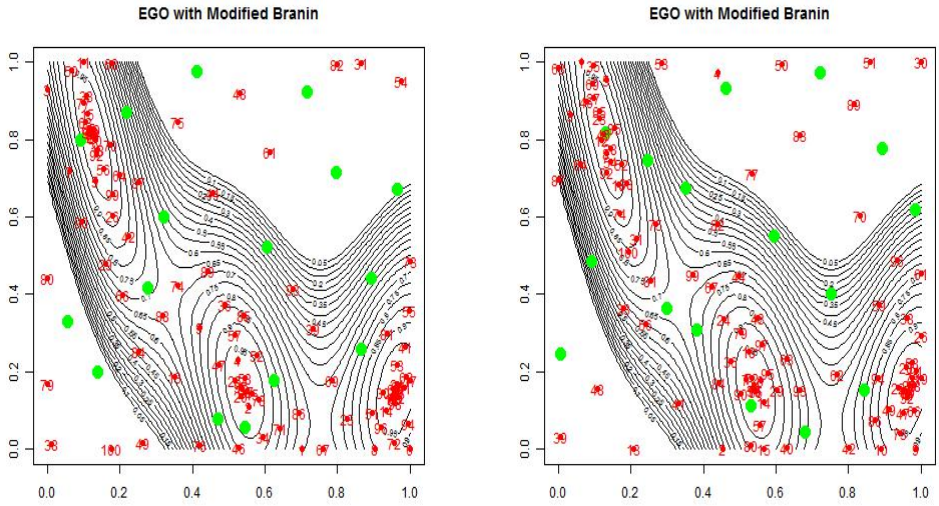

Figure 5.1: Examples of sampling sequence of EGO on modified Branin function. The green points are initial design points from Latin hypercube design. The red points are first 100 points in the sampling sequence.

5.4 The proposed sequential strategies

In Section 5.3, it was observed that directly using EGO does not generate a space-filling design within the feasible region. We now propose two model based sequential strategies, which are based on the following criteria: Intergrated MSE (IMSE) and Maximum MSE (MMSE).

5.4.1 IMSE and MMSE on feasible region

The MSE is a measure of the precision of predicted response at unsampled points. To ensure precision of estimation over the entire feasible region, we consider the mean square error integrated over the feasible region. Consider an N points fixed design $\{\mathbf{x}_i\}_{i=1}^N$ and the observed data $\mathbf{y}_N = (y_1, y_2, \cdots, y_N)$. Let the $N \times p$ matrix

$\mathbf{X}_N = \left(\mathbf{x}_1^T, \mathbf{x}_2^T, \cdots, \mathbf{x}_N^T\right)^T$ denote the matrix form of the sampled points $\{\mathbf{x}_i\}_{i=1}^N$. Then the *Integrated MSE*(IMSE) on the feasible region is defined by:

$$\begin{aligned} IMSE(\mathbf{X}_N, \mathbf{y}_N) &= \int_{f(\mathbf{x})\geq c} \sigma^2\left(\mathbf{x}|\mathbf{X}_N, \mathbf{y}_N\right) d\mathbf{x} \\ &= \int_{\Omega} \sigma^2\left(\mathbf{x}|\mathbf{X}_N, \mathbf{y}_N\right) I_{\{f(\mathbf{x})\geq c\}} d\mathbf{x}. \end{aligned} \tag{5.16}$$

where $I_{\{f(\mathbf{x})\geq c\}}$ is an indicator function which takes value 1 or 0 according as whether $\mathbf{x}$ lies in the feasible region.

Criterion $IMSE(\mathbf{X}_N, \mathbf{y}_N)$ is the integration of MSE of prediction over the feasible region. The notion of IMSE was first introduced by Sacks et al. (1989), in an attempt to minimize IMSE over the entire design space. IMSE can be considered as a measure of the overall uncertainty of prediction in the feasible region. Johnson et al. (1990) provided a theoretical justification of the IMSE criterion in terms of the A-optimality property.

The reason of using integration is that we want a design which leads to the same levels of uncertainty of prediction everywhere in the feasible region. Different points within the feasible region are treated equally. The integral in (5.16) can be evaluated by different numeric integration algorithms.

The *Maximum MSE* (MSE) over the feasible region is defined by:

$$MMSE(\mathbf{X}_N, \mathbf{y}_N) = \max_{\mathbf{x}\in\Omega} \sigma^2\left(\mathbf{x}|\mathbf{X}_N, \mathbf{y}_N\right) I_{\{f(\mathbf{x})\geq c\}}. \tag{5.17}$$

While IMSE measures the overall level of uncertainty of prediction in the feasible region, MMSE measures the maximum uncertainty of prediction over the feasible

region. Like the IMSE criterion, the MMSE criterion was also introduced and studied by Sacks et al. (1989).

It is impossible to directly optimize $IMSE(\mathbf{X}_N, \mathbf{y}_N)$ or $MMSE(\mathbf{X}_N, \mathbf{y}_N)$ over all possible fixed N points designs $\{\mathbf{x}_i\}_{i=1}^N$, since the boundary of feasible region is unknown. So a sequential strategy which can adaptively update the response surface will be considered. Similar to Chapter 4, we assume that the experiment has been conducted at $\{\mathbf{x}_i\}_{i=1}^n$, from which the response $\mathbf{y}$ has been observed. Two different strategies based on IMSE and MMSE will be proposed to select the next design point $\mathbf{x}_{n+1}$.

5.4.2 Using IMSE to generate sequential design

The first strategy is to select the next design point which can provide maximum improvement on IMSE. However, in the expression of IMSE in (5.16), the indicator function $I_{\{f(\mathbf{x})\geq c\}}$ is unknown. So we use the expectation of the indicator function, $\mathbb{E}\left(I_{\{f(\mathbf{x})\geq c\}}|\mathbf{X}_n, \mathbf{y}_n\right) = \mathbb{P}\left(f(\mathbf{x}) \geq c|\mathbf{X}_n, \mathbf{y}_n\right)$ to approximate $I_{\{f(\mathbf{x})\geq c\}}$. The quantity $\mathbb{P}\left(f(\mathbf{x}) \geq c|\mathbf{X}_n, \mathbf{y}_n\right)$, which is the probability that a given point is in the feasible region, can be obtained from (5.13). Consequently an approximate form of (5.16) is given by:

$$IMSE_1(\mathbf{X}_n, \mathbf{y}_n) = \int_\Omega \sigma^2\left(\mathbf{x}|\mathbf{X}_n, \mathbf{y}_n\right) \mathbb{P}\left(f(\mathbf{x}) \geq c|\mathbf{X}_n, \mathbf{y}_n\right) d\mathbf{x}. \tag{5.18}$$

Let $\mathbf{x}^*$ represent an unsampled point. Assume that if a new experiment is conducted at $\mathbf{x}^*$, response y^* will be observed. After conducting such an experiment,

$IMSE_1(\mathbf{X}_n, \mathbf{y}_n, \mathbf{x}^*, y^*)$ can be calculated as:

$$IMSE_1(\mathbf{X}_n, \mathbf{y}_n, \mathbf{x}^*, y^*) = \int_\Omega \sigma^2 \left(\mathbf{x}|\mathbf{X}_n, \mathbf{y}_n, \mathbf{x}^*, y^*\right) \mathbb{P}\left(f(\mathbf{x}) \geq c|\mathbf{X}_n, \mathbf{y}_n, \mathbf{x}^*, y^*\right) d\mathbf{x}. \tag{5.19}$$

Define the improvement on the precision of estimation in feasible region given by $\mathbf{x}^*, y^*$ as:

$$I_P(\mathbf{x}^*, y^*) = IMSE_1(\mathbf{X}_n, \mathbf{y}_n) - IMSE_1(\mathbf{X}_n, \mathbf{y}_n, \mathbf{x}^*, y^*). \tag{5.20}$$

$I_P(\mathbf{x}^*, y^*)$ measures the improvement after a new experiment is conducted and the response is obtained. The larger the value of $I_P(\mathbf{x}^*, y^*)$, the better is the prediction of the response surface in the feasible region at $\mathbf{x}^*, y^*$.

The improvement $I_P(\mathbf{x}^*, y^*)$ is a random variable since y^* is random. Similar to the EGO algorithm, the expected improvement, defined as the expectation of $I_P(\mathbf{x}^*, y^*)$ over y^* can be evaluated. The point with maximum expected improvement will be chosen at each iteration, i.e.:

$$\mathbf{x}_{n+1} = argmax_{\mathbf{x}^*} \mathbb{E}_{y^*} \left(I_P(\mathbf{x}^*, y^*)\right). \tag{5.21}$$

Since $IMSE_1(\mathbf{X}_n, \mathbf{y})$ is known from (5.20), the above algorithm is equivalent to:

$$\mathbf{x}_{n+1} = argmin_{\mathbf{x}^*} \mathbb{E}_{y^*} \left(\int_\Omega \sigma^2 \left(\mathbf{x}|\mathbf{X}_n, \mathbf{y}, \mathbf{x}^*, y^*\right) \mathbb{P}\left(f(\mathbf{x}) \geq c|\mathbf{X}_n, \mathbf{y}, \mathbf{x}^*, y^*\right) d\mathbf{x} \right). \tag{5.22}$$

Clearly, the expectation on the RHS of (5.21) is very difficult to compute, and unlike the EGO criterion in Jones et al. (1998), does not have a closed form solution. One possible method for evaluating the expectation is Monte Carlo simulation.

However, the uncertainties of the Monte Carlo simulation may lead to unsmooth improvement surface, which may lead to difficulty in search for the next points. A simpler strategy is to use the predicted value $\hat{y}(\mathbf{x})$ instead of $y(\mathbf{x})$ to avoid taking expectation over $y(\mathbf{x})$ in evaluating the improvement. Thus, we propose the following criterion to select the next point:

$$\mathbf{x}_{n+1} = argmax_{\mathbf{x}^*} I_P(\mathbf{x}^*, \hat{y}(\mathbf{x}^*)). \tag{5.23}$$

where the integral is evaluated through Quadrature rules.

Instead of calculating the expected improvement on IMSE, we assume that the actual response from $\mathbf{x}^*$ is the predicted value $\hat{y}(\mathbf{x}^*)$. This algorithm is equivalent to:

$$\mathbf{x}_{n+1} = argmin_{\mathbf{x}^*} \left(\int_{\Omega} \sigma^2 \left(\mathbf{x} | \mathbf{X}_n, \mathbf{y}_n, \mathbf{x}^*, \hat{y}(\mathbf{x}^*) \right) \mathbb{P} \left(f(\mathbf{x}) \geq c | \mathbf{X}_n, \mathbf{y}_n, \mathbf{x}^*, \hat{y}(\mathbf{x}^*) \right) d\mathbf{x} \right). \tag{5.24}$$

This approximation will overestimate improvement since it ignores uncertainty of a new observation. However, as long as it does not change the shape of the expected improvement surface over Ω, the approximation (5.23) is not expected to have significant influence on the sequential design.

To demonstrate this fact, a comparison of $I_P(\mathbf{x}^*, \hat{y}(\mathbf{x}^*))$ and $\mathbb{E}_{y^*} \left(I_P(\mathbf{x}^*, y^*) \right)$ over $\mathbf{x}^* \in \Omega$ is conducted using modified Branin function defined earlier in Section 4.7. We assume experiments have been conducted at a 25 points Latin hypercube design on $[0, 1]^2$. Simple Kriging method is used to model the response surface. Define grid G as:

$$G = \{ \mathbf{x} \in \Omega, \mathbf{x}_1 = 0.01 \times i, \mathbf{x}_2 = 0.01 \times j, 1 \leq i, j \leq 100 \} \tag{5.25}$$

For all unsampled points $\mathbf{x}^* \in G$, we would evaluate the improvement on $IMSE$ that $\mathbf{x}^*$ can provide. For $I_P(\mathbf{x}^*, \hat{y}(\mathbf{x}^*))$, the observation is assumed to be $\hat{y}(\mathbf{x}^*)$. For $\mathbb{E}_{y^*}\left(I_P(\mathbf{x}^*, y^*)\right)$, Monte Carlo simulation is used to evaluate the expectation. The plots of improvement for two different methods are shown in Figure 5.2. We find that even though the improvement in left plot is larger than improvement in right plot, their shape is very close to each other. So the approximation in (5.23) does not have significant influence on the sequential experiment design.

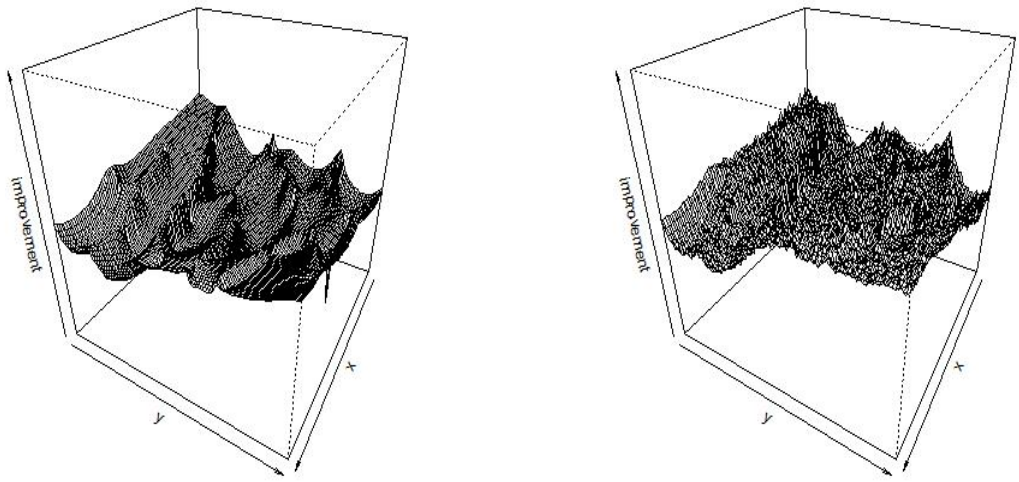

Figure 5.2: Evaluation of approximated expected improvement. The left plot evaluates $I_P(\mathbf{x}^*, \hat{y}(\mathbf{x}^*))$ and the right plot evaluates $\mathbb{E}_{y^*}\left(I_P(\mathbf{x}^*, y^*)\right)$.

5.4.3 Using MMSE to generate sequential design

The objective function (5.23) is still difficult to evaluate because of the integration over Ω. The computation power required is very intense when the dimension p is high.

A simpler algorithm is thus proposed as below:

$$\mathbf{x}_{n+1} = argmax_{\mathbf{x}^*}\sigma^2\left(\mathbf{x}^*|\mathbf{X}_n, \mathbf{y}_n\right)\mathbb{P}\left(f(\mathbf{x}^*) \geq c|\mathbf{X}_n, \mathbf{y}_n\right). \tag{5.26}$$

This algorithm directly selects the next point in the estimated feasible region with maximum prediction uncertainty. The term $\mathbb{P}\left(f(\mathbf{x}^*) \geq c|\mathbf{X}_n, \mathbf{y}_n\right)$ is used to carve out the infeasible region. The proposed criterion is an extension of the *Maximum MSE* (MMSE) criterion proposed by Sacks et al. (1989).

The above strategy will not directly optimize the expected improvement of IMSE over the feasible region. However, after the response from $\mathbf{x}_{n+1}$ is received, the MSE at $\mathbf{x}_{n+1}$, $\sigma^2(\mathbf{x}_{n+1}|\mathbf{X}_n, \mathbf{y}_n, \mathbf{x}_{n+1}, y(\mathbf{x}_{n+1}))$ will become 0. So it also provides improvement to the IMSE criterion given by (5.16).

Even though algorithm (5.26) cannot provide direct optimization to criterion function (5.16), it will require much less computational power. In the simulation part, both algorithms will be evaluated with different test functions.

5.4.4 Implementation

The implementation involves two stages: (i) Selection of the initial design and (ii) Choosing subsequent design points sequentially. Similar to Chapter 4, Latin Hypercube Design (LHD) is used as initial design. In each iteration of sequential design, the response surface is estimated using a Kriging predictor. We assume $q_m = 2, m = 1, \cdots, p$ where q_m is defined in (5.4). Then the point which can optimize (5.23) and (5.26) is selected. The sequential design is stopped when a pre-specified number of points is sampled. The two strategies will be referred to as GP-IMSE and

GP-MMSE hereforth.

5.5 Performance evaluation of the GP-IMSE and GP-MMSE

We now evaluate the performance of the GP-IMSE and GP-MMSE algorithms using simulations. The following performance measures are used to evaluate the sequential strategies:

1. P, RM, RA defined in Section 4.7.

2. The IMSE in the feasible region, $IMSE_f$, defined by:

$$IMSE_f = \mathbb{E}\Big[\int_{f(\mathbf{x})>c} \sigma^2\left(\mathbf{x}|\mathbf{x}_1, \cdots, \mathbf{x}_N, \mathbf{y}\right) d\mathbf{x}\Big]. \tag{5.27}$$

3. The MMSE in the feasible region, $MMSE_f$, defined by:

$$MMSE_f = \max_{f(\mathbf{x})>c} \{\sigma^2\left(\mathbf{x}|\mathbf{x}_1, \cdots, \mathbf{x}_N, \mathbf{y}\right)\}. \tag{5.28}$$

4. A measure of the 'average accuracy' of prediction within the feasible region, given by:

$$EF = \mathbb{E}\Big[\int_{f(\mathbf{x})>c} |\hat{y}(\mathbf{x}|\mathbf{x}_1, \cdots, \mathbf{x}_N, \mathbf{y}) - y(\mathbf{x})| d\mathbf{x}\Big]. \tag{5.29}$$

Both $IMSE_f$ and $MMSE_f$ measure the uncertainty within the feasible region. We want these values to be small. EF is also important, since it measures the prediction

accuracy of the design. Since both $IMSE_f$ and $MMSE_f$ depend on Gaussian process model, EF can provide a relative objective measurement on the prediction power of response surface within the feasible region.

5.5.1 Test with Branin function

Here, we will again use the modified Branin function given by (4.13) to evaluate the performances of GP-IMSE and GP-MMSE algorithms. A typical design with 100 runs generated by the GP-IMSE and GP-MMSE algorithms with threshold values $c = 0.2, 0.7$ are shown in Figure 5.3.

From the generated designs, , we find that both GP-IMSE and GP-MMSE perform significantly better than IDW-MM/IDW-MM-adj (see Figure 4.6) in carving out the infeasible region. The percentage of points in the feasible region is significantly higher, especially for the 0.7 threshold. Space-filling property within the feasible region also appears to be maintained. Moreover, like IDW-MM/IDW-MM-adj, both GP-IMSE and GP-MMSE are robust to the shape of feasible region.

Next, repeated simulations are conducted to evaluate the performance of GP-IMSE and GP-MMSE. The median performance measures computed from 500 repeated simulations are reported. The results are shown in Figures 5.4 and 5.5. In these figures, we also compare the IDW-MM, IDW-MM-adj and EGO algorithms to GP-IMSE/MMSE using the same performance measures.

From Figure 5.4 and 5.5, we find:

1. Both GP-IMSE and GP-MMSE can successfully carve out the infeasible region. The percentage of points in the feasible region after generating 100 design points

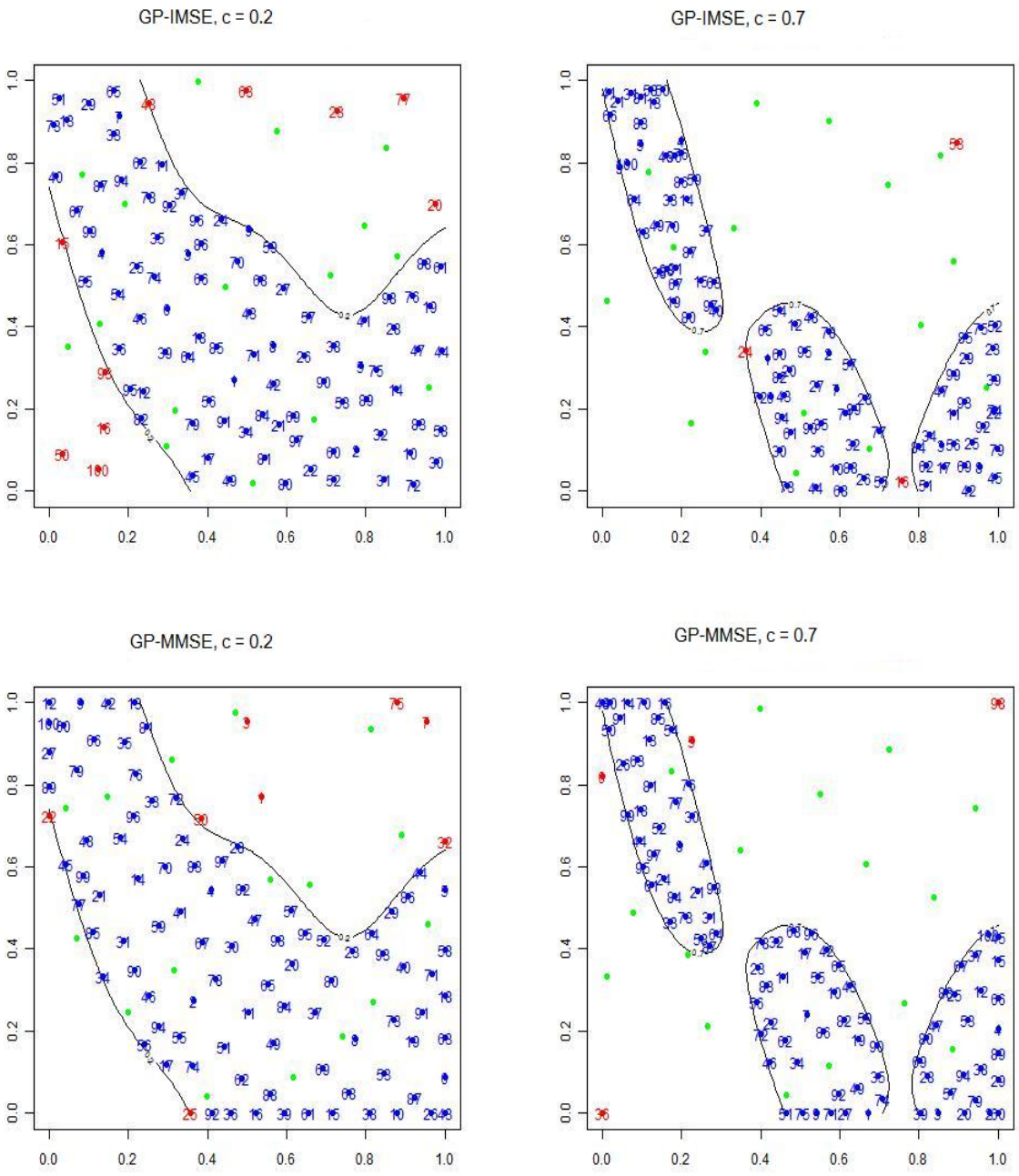

Figure 5.3: Space filling result by GP-IMSE and GP-MMSE for thresholds $c = 0.2, 0.7$.

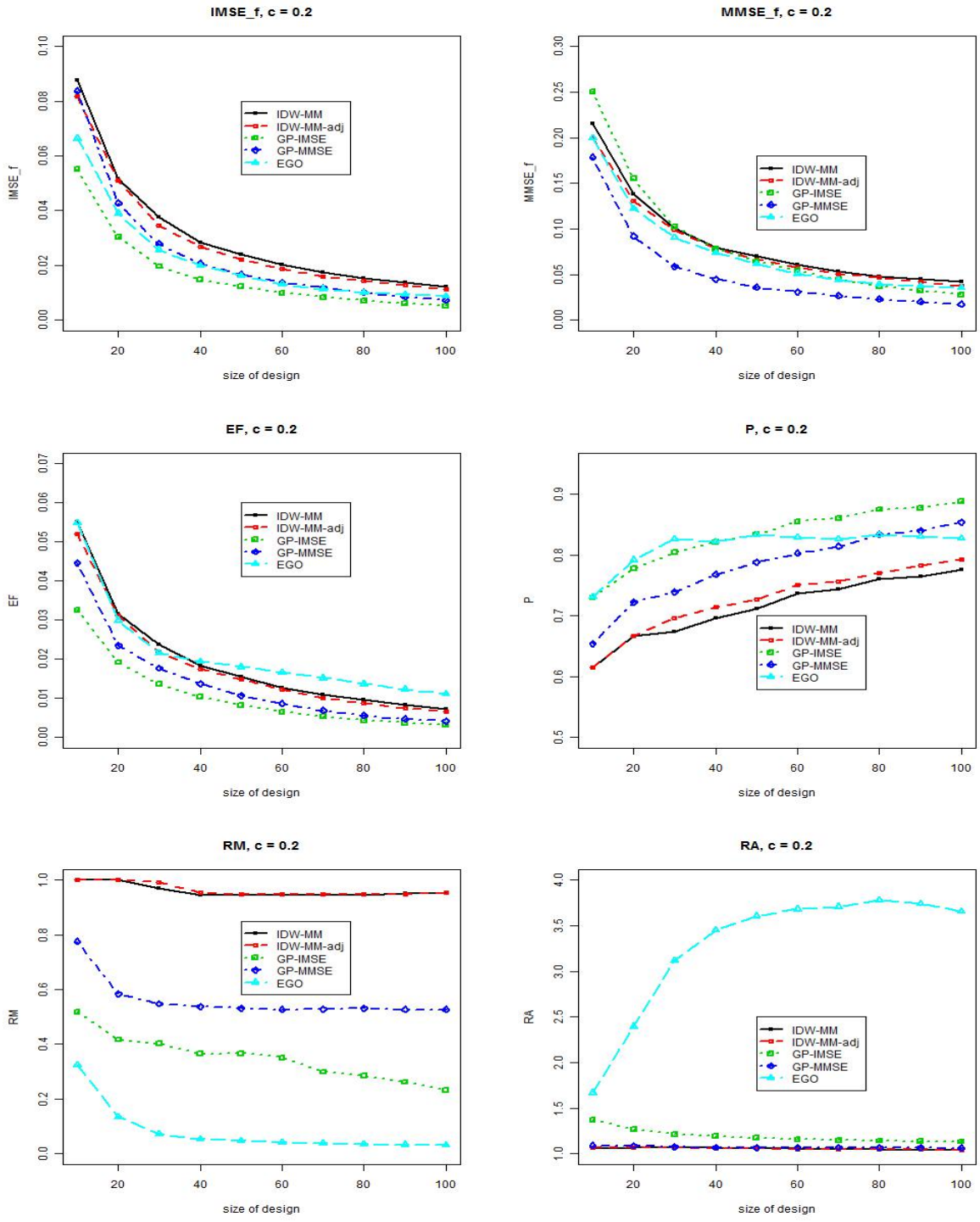

Figure 5.4: Performance of GP-IMSE/MMSE, $c = 0.2$.

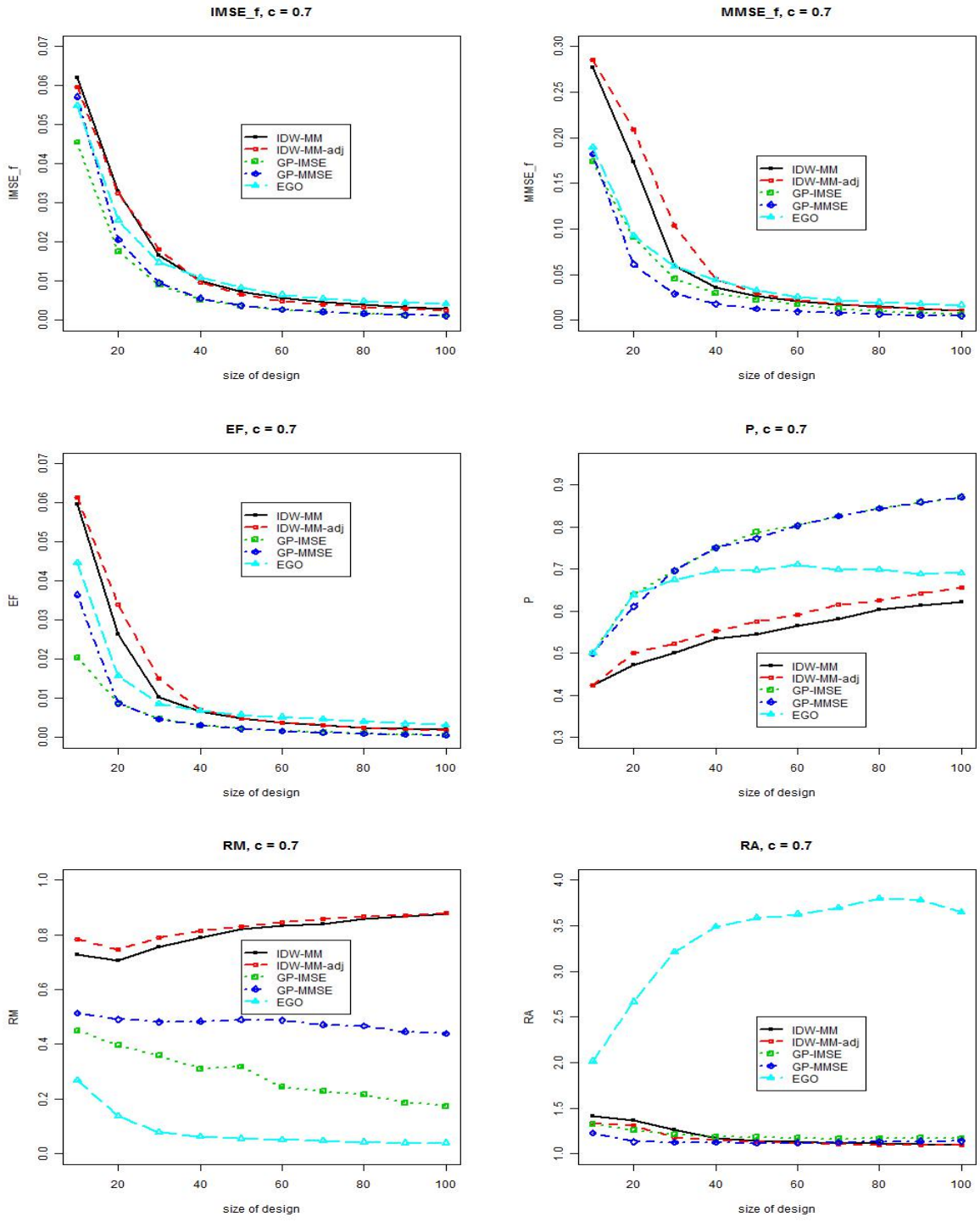

Figure 5.5: Performance of GP-IMSE/MMSE, $c = 0.7$.

is significantly higher than IDW-MM and IDW-MM-adj with optimal choice of k.

2. Performance measures $IMSE_f$, $MMSE_f$ and EF decrease when the sequential designs are generated in GP-IMSE and GP-MMSE. Both methods perform better than EGO.

3. Performance measures $IMSE_f$ is lower in GP-IMSE than GP-MMSE. On the contrary, $MMSE_f$ is lower in GP-MMSE than GP-IMSE. This is not surprising since $IMSE$ and $MMSE$ are the criteria for GP-IMSE and GP-MMSE respectively.

4. The GP-IMSE algorithm has the lowest EF, which means GP-IMSE can lead to most accurate estimation on unsampled points. We also observe that IDW-MM and IDW-MM-adj performs better than EGO.

5. The value of RM of GP-IMSE and GP-MMSE are lower than that of IDW-MM and IDW-MM-adj. On the contrary, RA of GP-IMSE and GP-MMSE are higher than IDW-MM and IDW-MM-adj. This means that the space-filling properties of GP-IMSE/MMSE are worse than IDW-MM/IDW-MM-adj. Moreover, EGO has the worst space-filling properties within the feasible region.

5.5.2 Preliminary test with modified Levy function

The space-filling design IDW-MM/IDW-MM-adj is based on grid-search. So it would be extremely hard to be applied to high dimensional response function. Since GP-IMSE/MMSE is not grid-based, it may be able to easily extend to high dimen-

sional response function. In this subsection, as a preliminary testing purpose, GP-MMSE is used to generate space-filling design with modified Levy function.

Levy function (Levy et al., 1981) is a widely used testing function in global optimization. The n dimension Levy function is defined as:

$$L(n, z_1, z_2, \cdots, z_n) = sin^2(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2 (1 + 10 sin^2(\pi y_i + 1)) + (y_n - 1)^2 (1 + sin^2(2\pi z_n)). \tag{5.30}$$

where $y_i = 1 + \frac{z_i - 1}{4}, i = 1, 2, \cdots, n$ and $|z_i| \leq 10$.

The n dimensional Levy function achieve global minimum at $(1, 1, \cdots, 1)^T$. The following modification can be used to adjust Levy function to $[0, 1]^n$:

$$L_1(n, x_1, x_2, \cdots, x_n) = L(n, 20x_1 - 10, 20x_2 - 10, \cdots, 20x_n - 10). \tag{5.31}$$

where $x_i \in [0, 1], i = 1, 2, \cdots, n$.

Finally the following modification is used to make the function similar to growth of nanowires:

$$f(n, x_1, x_2, \cdots, x_n) = \max\left(1 - \frac{L_1(n, x_1, x_2, \cdots, x_n)}{100}, 0\right) \tag{5.32}$$

GP-MMSE is used to generate space-filling design on the above modified Levy function with $n = 4$ and threshold $c = 0.4$. Since grid-based prediction on the response surface is required to evaluate the performance measures $IMSE_f$, $MMSE_f$, EF, RM and RA, only P will be reported in this section. The result can be found in Figure 5.6. The preliminary result can show GP-MMSE is successful in carving out

the infeasible region. Further evaluation would be conducted in future effort.

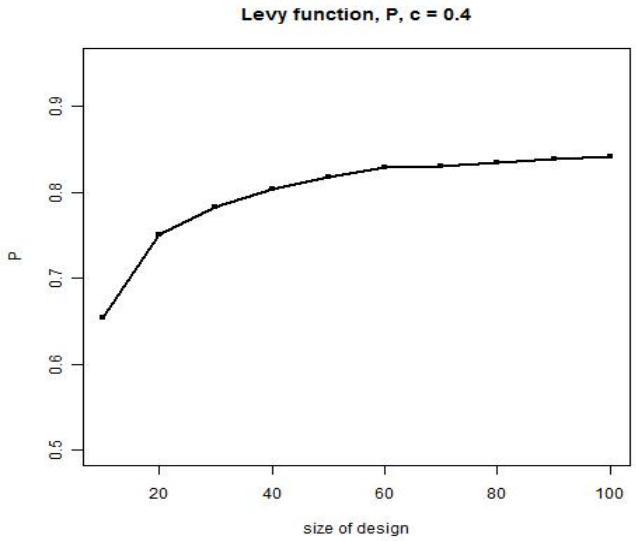

Figure 5.6: Performance of GP-MMSE on Levy function, $c = 0.4$.

5.6 Concluding remark

In this Chapter we have considered space-filling design with unknown constrains based on Gaussian process model. The objective of the design is to obtain precise estimation on the response surface of the feasible region. Estimation and prediction of Gaussian process model has been given and discussed. Then two sequential strategies which can enable precise estimation on response surface have been proposed and discussed. Intensive simulation studies have been used to evaluate the algorithm. Finally different extensions have been proposed for possible improvements on the sequential strategies.

Chapter 6

Extension of k-NN IDW to random response function and possible future work

In Chapters 4 and 5, two different types of strategies to generate space-filling design with unknown constraints on highly complex input-output response surface have been proposed and discussed. The proposed sequential strategies have been evaluated with simulation studies and have been seen to exhibit reasonably satisfactory performances. However, so far the proposed algorithm assumes that the response function $f(\mathbf{x})$ is deterministic. However, as in the motivating examples, in many situation the response is stochastic. For example, in Joseph et al. (2010), the growth of nanowires are estimated from r samples collected in each experiment. Then the observed yield $f(\mathbf{x})$ at design point $\mathbf{x}$ can be represented as $\frac{Y(\mathbf{x})}{r}$ where $Y(\mathbf{x})$ follows a binomial distribution with parameters r and $p(\mathbf{x})$, where $p(\mathbf{x})$ is the true value of

the response function. As an illustration, we assume that the true response function $p(\mathbf{x})$ is modified Branin function defined by (4.13). Then the observed response with different values of r is shown in Figure 6.1.

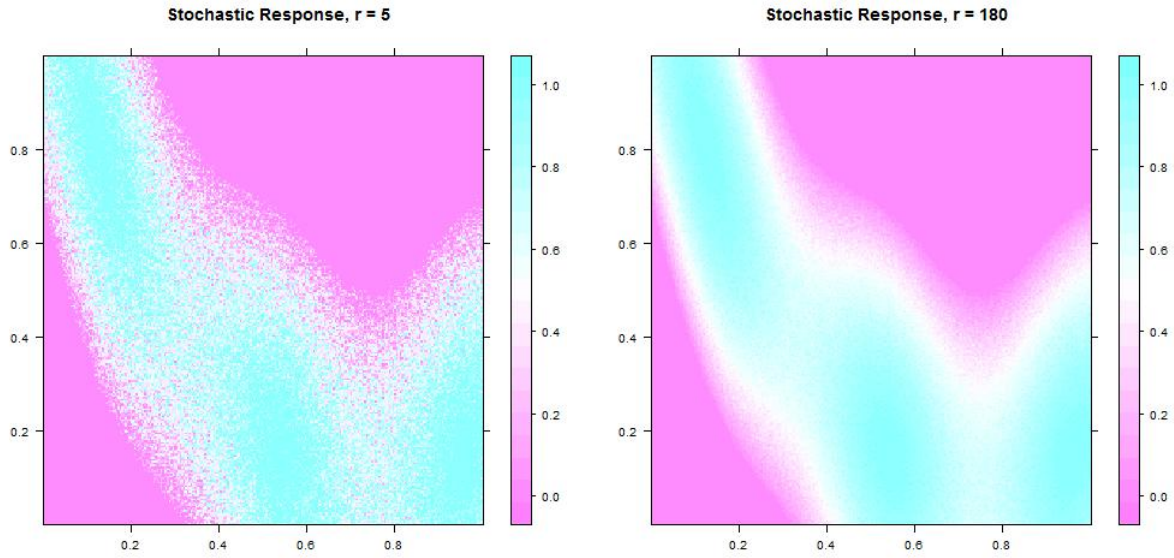

Figure 6.1: Example of contour plot for stochastic response for the modified Branin function. The sample size r is equal to 5 in left plot and 180 in right plot

As expected, when the sample size r is large, the observed response surface is similar to the true Branin function. However, when r is small, significant uncertainties are observed from the observation.

The randomness of the response makes the problem more complicated. The randomness introduces more uncertainties when we try to estimate the boundary and carve out the infeasible region. Moreover, when the response is random, most of the theoretical results derived in Chapter 4 do not hold.

6.1 Performance of IDW-MM on random response

The algorithms proposed in Chapter 4 and 5 can still be used to generate space-filling design on modified Branin function. Assume that the response represents a binomial proportion, s simulation framework is set up.

The true response function $p(\mathbf{x})$ is taken to be the modified Branin function in (4.13) and r is varied between 5 and 180. The same performance measures P, RM, RA defined in Section 4.7 are used to evaluate the algorithm. However, the expectation in the P, RM, RA are taken not only for the initial design but also over the distribution of the response function. The threshold c is assumed to be 0.2. Performance of IDW-MM with respect to different choice of r is shown in Figure 6.2.

From the simulation results shown in Figure 6.2, we find that the performance measure P of IDW-MM is worse than the deterministic case, especially when r is small. However, when r goes larger, the result will converge to the result in Figure 4.7. For IDW-MM, when k is large, the speed of convergence of P is fast since more data is used in estimation. However, we can also find performance measures RM and RA are robust to the choice of r.

6.2 Extension of IDW-MM for random response

When the response is random, the IDW-MM does not provide as accurate estimation as deterministic function. This is probably because, the weight function used in IDW-MM is not smooth. When a point to be estimated is very close to a sampled point, the weight given by that point will be much larger than other sampled points.

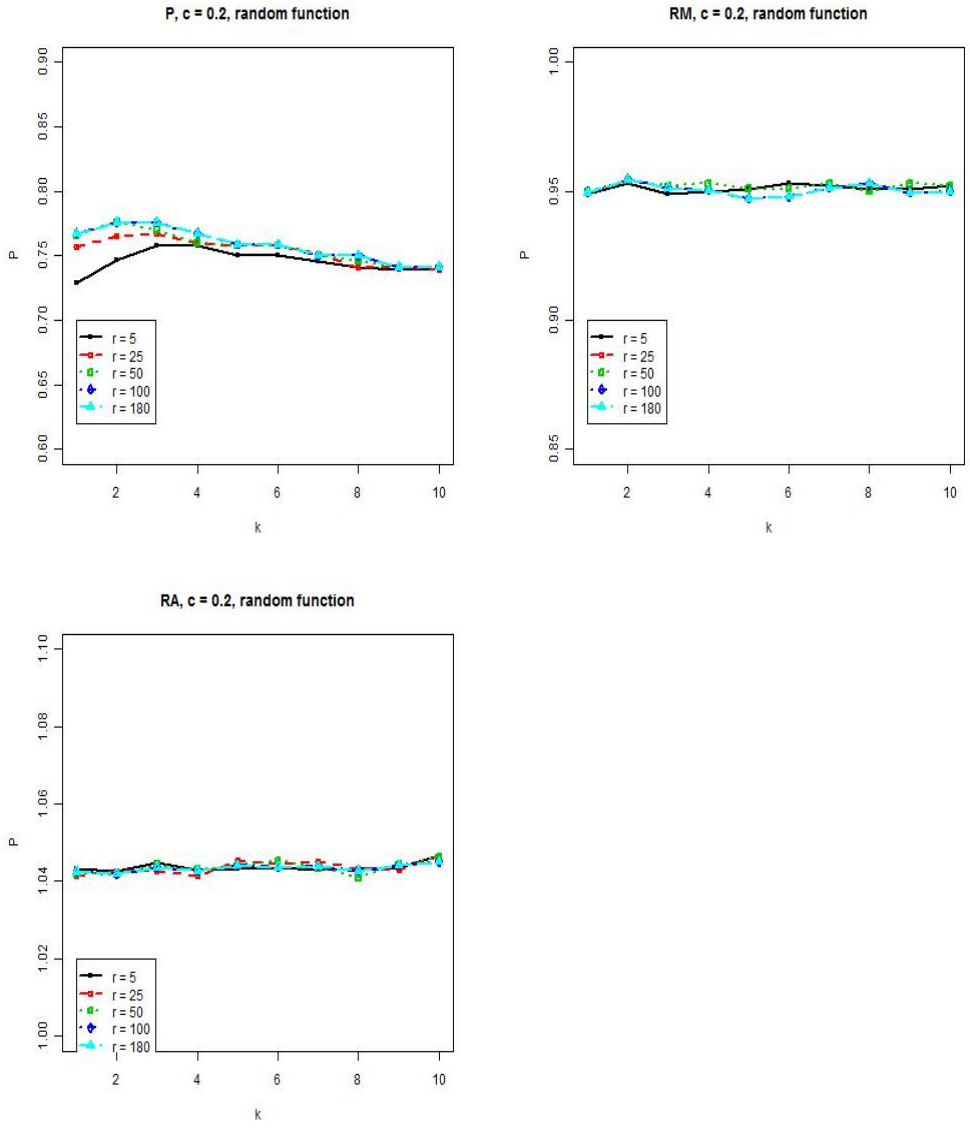

Figure 6.2: Performance of IDW-MM on random function with different r.

Consequently, the response from this sampled point will have an overwhelming impact on the estimation at the new point. Although this may not create a problem for a deterministic response, a random observation may lead to very poor estimates.

One possible improvement to the algorithm is to use more smoothing weights to estimate the response surface. Here a nonparametric regression approach inspired by Nadaraya-Watson (NW) kernel regression (Nadaraya, 1964) is explored to adjust the estimation.

Assume that n points $\mathbf{x}_1, ..., \mathbf{x}_n$ have been sampled. Then, the estimator based on NW weights (k-NN NW) for $\mathbf{x}$ will be:

$$\hat{f}_n(\mathbf{x}) = \frac{\sum_{i=1}^{n} I_k(\mathbf{x}_i, \mathbf{x}) w\left(\mathbf{x}, \mathbf{x}_i\right) f(\mathbf{x}_i)}{\sum_{i=1}^{n} I_k(\mathbf{x}_i, \mathbf{x}) w\left(\mathbf{x}, \mathbf{x}_i\right)}. \tag{6.1}$$

In (6.1), the weight function $w(\mathbf{x}, \mathbf{x}_i) = \left(1 - z^3\right)^3$, where $z = \frac{d(\mathbf{x}, \mathbf{x}_i)}{h}$, and h is the radius of the k-NN neighborhood of $\mathbf{x}$.

The above estimation is still based on k-NN framework. The theoretical results derived in Chapter 4 will still hold when k-NN NW is applied to deterministic functions. The advantage of NW weighting lies in the fact that the weight function $w(\mathbf{x}, \mathbf{x}_i)$ is much more smooth than the one used in IDW weighting. This aspect allows the method to incorporate information from more points into estimation, which is important when the response is random. It is worthwhile to note that, k-NN NW is no longer an interpolation method, since the estimated values at sampled points are no longer the same as observed values. However, this make sense since the response is actually random.

We now use the k-NN NW to generate space filling design (referred as NW-MM)

on modified Branin function with $c = 0.2$. The results are shown in Figure 6.3, from which the following conclusions can be obtained:

1. The performance measures P are more robust to the choice of k than IDW-MM.

2. The smoothing weights given by k-NN NW can lead to space filling design much more robust to the choice of r with respect to performance measure P. The difference between results obtained from $r = 5$ is much more similar with the result obtained from $r = 180$. Satisfactory result can be obtained even when $r = 5$.

3. The optimal k in NW-MM is larger than IDW-MM and IDW-MM-adj due to the smoothing weights. Moreover, the value of P in NW-MM with optimal k (0.787) is better than P obtained from IDW-MM with optimal k (0.776).

4. The performance measure RM and RV is similar to algorithm with IDW-MM. Moreover, the values of RM and RV still do not depend on the choice of k or r.

6.3 Other possible extensions and future work

There are several other ways we can possibly improve the proposed strategies:

1. Improving the precision of estimation: Both IDW and simple Kriging can be improved by adding regression terms into the estimation. For example, IDW with regression terms has been described in Joseph and Kang (2011). Universal Kriging (Santner et al., 2003) and Blind Kriging (Joseph and Hung, 2008) are

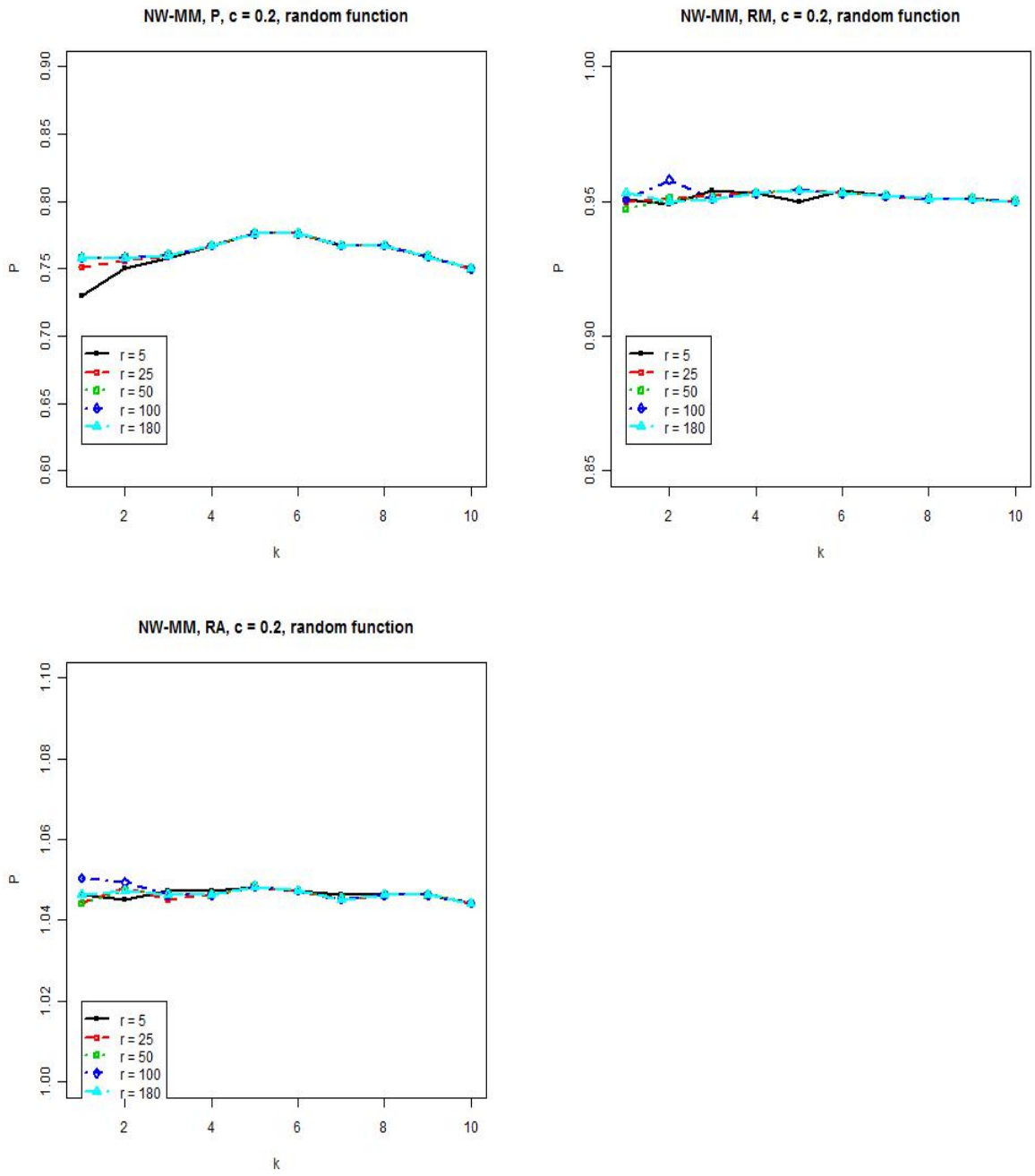

Figure 6.3: Performance of IDW-MM on random function with different r.

extensions of simple Kriging in which the mean part of Gaussian process model given by (5.2) is expressed as a linear combination of functions, i.e., $\mu(\mathbf{x}) = \sum_{j=1}^{n} \beta_j f_j(\mathbf{x})$.

2. Increase the efficiency of optimization algorithms. The current version of IDW-MM algorithm is based on a grid-search strategy. Possible optimization algorithms which can avoid the grid framework may lead to more computational efficient designs. Moreover, the possibility of improving the optimization algorithm for GP-IMSE/MMSE also needs be explored to speed up the computation.

Bibliography

J. Atherton, B. Charbonneau, D. B. Wolfson, L. Joseph, X. Zhou, and A. C. Vandal. Bayesian Optimal Design for Changepoint Problems. *Canadian Journal of Statistics*, 37:495 – 512, 2009.

S. Ba and V. R. Joseph. Multi-Layer Designs for Computer Experiments. *Journal of the American Statistical Association - Theory and Methods*, 106(495):1139 – 1149, 2011.

C. Barancourt, J.D. Creutin, and J. Rivoirard. A method for delineating and estimating rainfall fields. *Water resources research*, 28(4):1133–1144, 1992.

S. Biedermann, H. Dette, and D. C. Woods. Optimal Designs for multivariable spline models. monograph available at http://eprints.soton, ac.uk/69157, 2009.

G. E. P. Box and W. G. Hunter. Sequential Design of Experiments for Nonlinear Models. *Proceedings of the IBM Scientific Computing Symposium on Statistics*, pages 113 – 137, 1965.

F. H. Branin. Convergent Method for Finding Multiple Solutions of Simultaneous Nonlinear Equations. *IBM J. Res. Develop*, 16(4):113 – 137, 1972.

D. Bremner, E. Demaine, J. Erickson, J. Iacono, S. Langerman, P. Morin, and G. Toussaint. Output-sensitive algorithms for computing nearest-neighbor decision boundaries. *Discrete and Computational Geometry*, 33(4):593–604, 2005.

A.D. Bull. Convergence Rates of Efcient Global Optimization Algorithms . *Journal of Machine Learning Research*, 12:2879–2904, 2011.

P. Chaudhuri and P. A. Mykland. Nonlinear Experiments: Optimal design and Inference Based on Likelihood. *Journal of the American Statistical Association*, 88 (422):538 – 546, 1993.

H. Chernoff. Locally Optimal Designs for Estimating Parameters. *Annals of Mathematical Statistics*, 24(4):586 – 602, 1953.

H. Chernoff. Approaches in Sequential Design of Experiments. Technical reptort, 1973.

T. Dasgupta. *Robust parameter design for automatically controlled systems and nanostructure synthesis*. PhD thesis, Georgia Institute of Technology, 2007.

T. Dasgupta, C. Ma, R. Joseph, Z. L. Wang, and C. F. J. Wu. Statistical Modeling and Analysis for Robust Synthesis of Nanostructures. *Journal of the American Statistical Association*, 103(482):594 – 603, 2008.

H.A. Dror and D. M. Steinberg. Sequential Experimental Designs for generalized Linear Models. *Journal of the American Statistical Association*, 103(481):288 – 298, 2008.

K.T. Fang and D.K.J. Lin. *Uniform experimental designs and their applications in industry, Handbook of Statistics*. North Holland, Amsterdam, 2000.

V. V. Federov. *Theory of Optimal Experiments*. Academic Press, New York, 1972.

I. Ford and S. D. Silvey. A Sequentially Constructed Design for Estimating a Nonlinear Parametric Function. *Biometrika*, 67(2):381 – 388, 1980.

R. Franke. A Critical Comparison of Some Methods for Interpolation of Scattered Data. Technical Report, 1989.

A.E. Gelman and X.-L. Meng. A Note on Bivariate Distributions that are Conditionally Normal. *The American Statistician*, 45:125–126, 1991.

L. M. Haines. Optimal Designs for Nonlinear regression Models. *Communications in Statistics, Part A - Theory and Methods*, 22(6):1613 – 1627, 1993.

N. Henkenjohann, R. Gobel, M. Kleiner, and J. Kunert. An Adaptive Sequential Procedure for Efficient Optimization of the Sheet Metal Spinning Process. *Quality and Reliability Engineering International*, 21(5):439 – 455, 2005.

G. Hohmann and W. Jung. On Sequential and Nonsequential D-Optimal Experiment Design. *Biometrisch Zeitschrift*, 17(5):329 – 336, 1975.

Q. Huang, L. Wang, T. Dasgupta, L. Zhu, P. K. Sekhar, S. Bhansali, and Y. An. Statistical Weight Kinetics Modeling and Estimation for Silica Nanowire Growth Catalyzed by Pd Thin Film. *IEEE Transactions on Automation Sciences and Engineering*, 8(2):303–310, 2011.

M.E. Johnson, L.M. Moore, and D. Ylvisaker. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2):131–148, 1990.

D.R. Jones. A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization*, 21(4):345–383, 2001.

D.R. Jones, M. Schonlau, and WJ. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

R. Joseph. Limit Kriging. *Technometrics*, 48(4):458–466, 2006.

R. Joseph and Y. Hung. Orthogonal-maximin Latin hypercube design. *Statistica Sinica*, 18(1):171–186, 2008.

R. Joseph and L. Kang. Regression-based inverse distance weighting with applications to computer experiments. *Technometrics*, 53(3):254–265, 2011.

R. Joseph, T. Dasgupta, and C. F. J. Wu. Minimum Energy Designs: From Nanostructure Synthesis to Sequential Optimization. under revision, 2010.

A. Jourdana and J. Francob. A new criterion based on Kullback-Leibler information for space filling designs. under revision, 2006.

S. Karlin and W. J. Studden, editors. *Tchebycheff Systems, with Applications in Analysis and Statistics*. Interscience Publishers, New York, 1966.

D. Lazzaro and L. B. Montefusco. Radial Basis Functions for the Multivariate Interpolation of Large Scattered Data Sets. *Journal of Computational and Applied Mathematics*, 140(1):521–536, 2002.

AV. Levy, A. Montalvo, S. Gomez, and A. Calderon. Topics in global optimization. Lecture Notes in Mathematics No. 909, 1981.

J. Liu, editor. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 2002.

J.C. Lu, S. L. Jeng, and K. Wang. A Review of Statistical Methods for Quality Improvement and Control in Nanotechnology. *Journal of Quality Technology*, 41 (2):148–164, 2009.

S. Lukaszyk. A New Concept of Probability Metric and Its Applications in Approximation of Scattered Data Sets. *Computational Mechanics*, 33(4):299–304, 2004.

M.D. McKay, R.J. Beckman, and W.J. Conover. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239–245, 1979.

T.J. Mitchell, M.D. Morris, and D. Ylvisaker. Two-Level Fractional Factorials and Bayesian Prediction. *Statistica Sinica*, 5(2):559–573, 1995.

M.D. Morris and T.J. Mitchell. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43(3):381–402, 1995.

S. Mukhopadhyay and L. M. Haines. Bayesian D-optimal Designs for the Exponential Growth Model. *Journal of Statistical Planning and Inference*, 44(3):385–397, 1995.

E. A. Nadaraya. On Estimating Regression. *Theory of Probability and its Applications*, 9(1):141–142, 1964.

R.R. Picard and R.D. Cook. Cross-Validation of Regression Models. *Journal of the American Statistical Association*, 79(387):575–583, 1984.

C. Pozrikidis. *Numerical Computation in Science and Engineering*. Oxford Univ Press, Oxford, 1998.

P. Z. G. Qian, H. Wu, and C. F. J. Wu. Gaussian Process Models for Computer Experiments with Qualitative and Quantitative Factors. *Technometrics*, 50(3): 383–396, 2008.

P. Ranjan, D. Bingham, and G. Michailidis. Sequential Experiment Design for Contour Estimation from Complex Computer Codes. *Technometrics*, 50(4):527–541, 2008.

A. Roy, S. Ghosal, and W. F. Rosenberger. Convergence properties of sequential Bayesian D-optimal designs. *Journal of Statistical Planning and Inference*, 139(2): 425–440, 2009.

J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and Analysis of Computer Experiments. *Statistical Sciences*, 4(4):409–435, 1989.

T. J. Santner, B. J. Williams, and W. I. Notz. *The Design and Analysis of Computer Experiments*. Springer, New York, 2003.

D. Shepard. A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 ACM National Conference*, pages 517–524, 1968.

R. Sibson. Contribution to Discussion of papers by H. P. Wynn and P. J. Laycock. *Journal of the Royal Statistical Society, Ser. B*, 34:181–183, 1972.

S. D. Silvey. Contribution to Discussion of papers by H. P. Wynn and P. J. Laycock. *Journal of the Royal Statistical Society, Ser. B*, 34:174–175, 1972.

S. D. Silvey. *Optimal Design*. Chapman and Hall, London, 1980.

S. D. Silvey and D. M. Titterington. A Geometric Approach to Optimal Design Theory. *Biometrika*, 60(1):21–32, 1973.

E. Stinstra, D. Den Hertog, P. Stehouwer, and A. Vestjens. Constrained Maximin Designs for Computer Experiments. *Technometrics*, 45(4):340–346, 2003.

B. Tang. Orthogonal Array-Based Latin Hypercubes. *Journal of the American Statistical Association*, 88(424):1392–1397, 1993.

D. M. Titterington. Optimal Design: Some Geometrical Aspects of D-Optimality. *Biometrika*, 62(2):313–320, 1975.

L. Vandenberghe, S. Boyd, and S. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19: 499–533, 1998.

C. F. Wu and M. Hamada. *Experiments: Planning, Analysis, and Parameter Design Optimization*. Wiley, 2000.

Z. Yao and W.L. Ruzzo. A regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data. *BMC Bioinformatics*, 7(Suppl 1): S11, 2006.